



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Matemáticas

Escuela Profesional de Estadística

Aplicación de la técnica multivariante de escalamiento multidimensional en el Programa Internacional de Evaluación de Estudiantes (PISA)

TESINA

Para optar el Título Profesional de Licenciada en Estadística

AUTOR

Yuri SÁNCHEZ ACOSTUPA

Lima, Perú

2016



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Sánchez, Y. (2016). *Aplicación de la técnica multivariante de escalamiento multidimensional en el Programa Internacional de Evaluación de Estudiantes (PISA)*. [Tesina de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, Escuela Profesional de Estadística]. Repositorio institucional Cybertesis UNMSM.



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

Universidad del Perú, DECANA DE AMÉRICA

FACULTAD DE CIENCIAS MATEMÁTICAS

PROGRAMA DE ACTUALIZACIÓN PARA LA TITULACIÓN PROFESIONAL 2016 - II
MODALIDAD EXAMEN DE SUFICIENCIA PROFESIONAL
(RR. N° 03849-R-16)

ESCUELA PROFESIONAL DE ESTADÍSTICA

ACTA DE EXPOSICIÓN DE TESINA

En la Ciudad Universitaria, Facultad de Ciencias Matemáticas, siendo las 10:30 horas, del día 12 de diciembre del 2016, se reunieron las docentes designadas como miembros del Jurado Evaluador:

- | | |
|--|------------|
| - Dra. Ilse Janine Villavicencio Ramírez | Presidenta |
| - Lic. Zoraida Huamán Gutiérrez | Miembro |

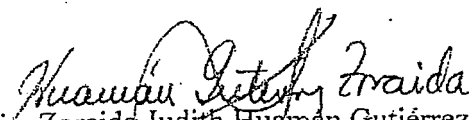
Para la exposición de la Tesina titulada: «*Aplicación de la Técnica Multivariante de Escalamiento Multidimensional en el Programa Internacional de Evaluación de Estudiantes (PISA)*» presentado por el Bachiller Yuri Sánchez Acostupa.

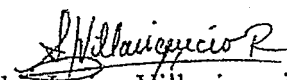
Luego de la exposición de la tesina, los Miembros del Jurado hicieron las preguntas correspondientes, a las cuales el Bachiller Yuri Sánchez Acostupa respondió con acierto y solvencia, demostrando pleno conocimiento del tema.

Hecha la evaluación correspondiente, según tabla adjunta, el Bachiller Yuri Sánchez Acostupa Ibeth mereció la aprobación obteniendo como calificativo promedio y la nota de dieciocho 18 (letras y números).

A continuación los Miembros del Jurado Evaluador, dan manifiesto que el Bachiller Yuri Sánchez Acostupa, aprobó la exposición de la Tesina.

Siendo las 11:00 horas, se levantó la sesión, firmando para constancia la presente acta en dos (2) copias originales.


Lic. Zoraida Judith Huamán Gutiérrez
MIEMBRO


Dra. Ilse Janine Villavicencio Ramírez
PRESIDENTA

Dedicatoria

A Dios, porque en su infinita gracia me concede la vida.

A mis padres Delfina, Tomas y hermanos,
por su paciencia y entendimiento hacia mi persona, quienes me
apoyaron y alentaron a poder trazar mis objetivos personales.

Agradecimiento

Debo agradecer a mis Profesoras de las Escuela Académica
Profesional de Estadística por brindarme su paciencia y perseverancia
en mi formación como profesional y como persona, quienes me
motivaron para el avance, de mí trabajo de investigación.

Resumen

En la presente investigación se pretende aplicar la técnica de escalamiento multidimensional (EM), como una alternativa para poder identificar el comportamiento de los datos objetivamente y en la determinación del posicionamiento de los puntos en el espacio d-dimensional resultante para poder establecer la similitud y diferencias.

Estableciendo como objetivo principal determinar los factores que caracterizan a las Institución Educativa que influyen en la evaluación del aprendizaje en el Programa Internacional de Evaluación de Estudiantes (PISA) aplicado en el país en el año 2012.

Palabras Claves: Escalamiento multidimensional (EM), Programa Internacional de Evaluación de Estudiantes (PISA).

Summary

he presentation of the research the multidimensional scaling technique (EM) was applied as an alternative to be able to identify the behavior of the data objectively and in the determination of the positioning of the points in the resulting d-dimensional space to be able to establish The similarity and differences.

Establishing as main objective to determine the factors that characterize the Educational Institution that influenced the evaluation of learning in the International Program of Student Assessment (PISA) applied in the country in 2012.

Key Words: Multidimensional Scaling (MS), International Student Assessment Program (PISA).

Índice

Contenido

Dedicatoria.....	2
Agradecimiento.....	2
Resumen	3
Summary.....	3
Índice.....	5
Capítulo I.....	9
Introducción	9
1.1.- Situación Problemática	9
1.2.- Formulación del Problema	9
1.3.- Objetivos	10
1.3.1.- Objetivos Generales	10
1.4.- Justificación y Originalidad	10
1.4.1.- Originalidad	10
1.4.2.- Justificación	10
Capítulo II:	11
Marco Teórico	11
2.1.- Antecedentes	11
2.1.1. - Antecedentes Nacionales	11
2.1.2.- Antecedentes Internacionales.....	13
2.2.- Bases Teóricas Escalamiento Multidimensional (EM)	16
2.2.1.- Reseña Histórica	16
2.2.2.-Conceptos.....	18
a.- Medidas de Similitud	18
b.- Coeficientes de asociación	19
c.-Medidas de distancia	19
d.- Medidas de proximidad y de Distancia	20
e.-Tipos de datos.....	20
2.3.3.- Medidas de proximidad.....	21
f.-Medidas para Variables Cuantitativas	21
g.- Coeficiente de Congruencia	21
h.-Coeficiente de Correlación.....	21

i.-Medidas para Datos Binarios	22
2.3.5.- Modelo General EM.....	25
j.- Modelo de EM Métrico.....	25
l.- Escalamiento Clásico	25
m.- Distancia Euclídea.	28
n.- Distancia de Minkowski.....	28
ñ.- Distancia de Mahalanobis	29
2.3.6.- Recuperación de Coordenadas	29
2.3.7.- EM No Métrico.	33
2.3.8.- Número de Dimensiones	36
Capítulo III:	38
Metodología	38
3.1.- Tipo y Diseño de Investigación	38
3.2.- Unidad de Análisis	38
3.3.- Población de Estudio.....	38
3.4.-Tamaño de Muestra.....	38
3.5.- Selección de Muestra	38
3.6.- Técnica de Recolección de Datos	39
3.7.- Mapas Perceptuales.....	39
3.10.-El Programa Internacional para la Evaluación de Estudiantes (PISA) de la Organización para la Cooperación y el Desarrollo Económico (OCDE)	40
Capítulo IV:.....	43
Resultados y Discusión	43
Aplicación Del EM No Métrico	43
4.1.- Análisis, interpretación y discusión de resultados	43
4.1.1.-Selección de variables de la Base de datos PISA.....	43
4.2.-Resultados del Análisis EM No Métrico en SPSS	46
CONCLUSIONES	53
BIBLIOGRAFÍA:.....	54
ANEXOS I	56
ANEXOS II	62

Índice de Tablas

Tabla N° 1: Tabla de Contingencia para objetos r y s	22
Tabla N°2: Medidas de Similaridad Datos Binarios	23
Tabla N° 3.-Variables Seleccionadas para el Análisis EM No Métrico	45
Tabla N° 4.- Medidas de Bondad Ajuste del Modelo	46
Tabla N° 5 .Valor de Stress y RSQ	46
Tabla N° 6. Medidas de ajuste y stress.....	49
Tabla N° 7. Coordenas en la Configuración.....	50

Índice de Figuras

Figura N°1: Valor del Stress según el número de Dimensiones.	37
Figura N°2: Selección de Variables de la Base de Datos PISA.	44
Figura N°3: Selección de Variables de la Base de Datos PISA.	44
Figura N°4: Gráfico de Ajuste Lineal.	47
Figura N°5: Gráfico de Ajuste No Lineal.	48
Figura N°6: Grafico Shepard.....	48
Figura N°7. Mapa Perceptual de Los Factores de la I.E.	51

Capítulo I

Introducción

1.1.- Situación Problemática

La Educación es uno de los pilares para el desarrollo del país y que determina el desarrollo social; con el cual resulta posible acceder a mejores oportunidades laborales, velar por el bienestar del presente y futuro de las familias, por la cual la educación recobra una especial importancia en países con elevados niveles de desigualdad y pobreza, como es el caso de nuestro país, conocer la realidad en que se encuentran las instituciones educativas del país, puede determinar el futuro educativo de los estudiantes, por lo que se requieren poner énfasis en la educación.

El Perú está creciendo económicamente, pero ¿quiénes son los que realmente se benefician con este crecimiento? ¿Podemos exigir más a nuestros estudiantes y maestros un país que exigentemente destina menos del 3% del PBI a educación? (Victormazzi, 2012)

1.2.- Formulación del Problema

¿Qué factores caracterizan a las Instituciones Educativas y se relacionan en la evaluación del aprendizaje en el Programa Internacional de Evaluación de Estudiantes (PISA) del año 2012?

1.3.- Objetivos

1.3.1.- Objetivos Generales

Determinar los factores caracterizan a de las Instituciones Educativas y se relacionan en la evaluación del aprendizaje en el Programa Internacional de Evaluación de Estudiantes (PISA) del año 2012, mediante la Técnica Multivariante de Escalamiento Multidimensional (EM).

1.4.- Justificación y Originalidad

1.4.1.- Originalidad

Aquí en el Perú y en el mundo, no existen estudios sobre la aplicación de la técnica multivariante de Escalamiento Multidimensional (EM) en la evaluación de estudiantes del Programa Internacional de Evaluación de Estudiantes (PISA).

Establecer las similitudes y diferencias del programa de evaluación Programa Internacional de Evaluación de Estudiantes (PISA), a través del Escalamiento Multidimensional EM, constituye una herramienta muy útil para poder determinar si hay una relación entre las variables.

1.4.2.- Justificación

La importancia de este problema se basa en emplear la técnica estadística de Escalamiento multidimensional (EM), nos va permitir determinar y conocer el posicionamiento de las Instituciones educativas, para poder agruparlas de acuerdo a las características con cuenta para afrontar la enseñanza educativa con que brindan a los estudiantes.

Capítulo II:

Marco Teórico

En el presente capítulo se divisarán conceptos básicos desde el punto de vista teórico; que son el soporte para la presente investigación sobre la técnica escalamiento multidimensional (EM) de análisis multivariado.

2.1.- Antecedentes

2.1.1. - Antecedentes Nacionales

Salazar (2006), realiza una investigación sobre : *Análisis Procrustes: Comparación del Escalamiento Multidimensional y Análisis de Correspondencia*, UNMSM-PERU, con el objetivo principal de hacer una comparación de las técnicas Multivariantes Escalamiento Multidimensional y Análisis de Correspondencia, a través del Análisis Procrustes , para determinar el posicionamiento de las ciudades del interior del país y su asociación con las ramas de la actividad económica más importantes haciendo el uso de los mapas perceptuales con cada una de ellas , se utilizó los datos del Ministerio de Trabajo y Promoción del Empleo recogidos a través de la Dirección nacional de Promoción del Empleo y Formación Profesional obtenidos de la Encuesta de Hogares Especializada de Niveles de Empleo(ENAHU).

Los resultados que se obtuvieron en la aplicación es que existen diferencias. El análisis de escalamiento multidimensional permite la comparación entre objeto y/o entidades (ciudades), mientras que el Análisis de Correspondencia aporta

un medio de comparación directa de similitud y diferencias de ciudades asociados con las ramas de la actividad, condicionada por un conjunto de atributos, el mapa perceptual es una técnica que ofrece comparaciones conjuntas que no está al alcance de otro método multivariante. El Análisis Procrustes hace la comparación mediante una matriz de contingencia.

Ucedo (2013), presenta una investigación: Comparación De Los Modelos Logit Y Probit Del Análisis Multinivel, En El Estudio Del Rendimiento Escolar, UNMSM-PERU, donde aplica los modelos de multinivel. Donde menciona que es la metodología de análisis más adecuada para explicar datos jerarquizados, hace una comparación entre dos modelos Logit y el Probit para describir la relación existente entre una variable de respuesta de tipo dicotómica y una o más variables explicativas.

El estudio hace una comparación de los modelos de análisis multinivel con el objetivo conocer las similitudes, diferencias y conocer las ventajas y desventajas en la aplicación de cada uno de los modelos. Para el análisis empleó la Base de Datos del Programa Internacional para la Evaluación de Estudiantes (PISA) del año 2009.

Estableció que al incrementarse el número de variables en cada uno de los modelos del análisis multinivel a través de sus enlaces Logit y Probit la varianza estimada como la desvianza tiende a reducirse, se puede señalar en cuanto a su función presentó un buen ajuste a diferencia del ajuste global donde el modelo Probit; es el que presentó un mejor ajuste en ambos casos el tamaño de muestra tiene que ser el mismo.

Recomienda tener un cuidado especial; si se tratan de diferencias debido a que manejan funciones de enlaces matemáticamente diferentes por lo cual las estimaciones no coincidan.

2.1.2.- Antecedentes Internacionales

Guerrero & Ramírez (2002) presenta un artículo: *El Análisis De Escalamiento Multidimensional: Una Alternativa Y Un Complemento A Otras Técnicas Multivariantes*, Universidad Pablo de Olavide-ESPAÑA, donde su objetivo es poder hacer una comparación de las técnicas Multivariantes análisis de Escalamiento Multidimensional con el Análisis Factorial y Clúster, ante la proliferación de datos y el fácil acceso a los mismos, haciendo la aplicación para poder comparar a través de los datos que ha obtenido de la Encuesta de Coyuntura Turística de Andalucía (ECTA) y de la Encuesta de Ocupación Hotelera de la Junta de Andalucía pretendiendo analizar la infraestructura turística de Andalucía con el objetivo de identificar aquellas ciudades que sean más similares.

Se llegó a identificar dos dimensiones que la caracterizaban a las ciudades turísticas de Málaga, Sevilla, Cádiz y Granada las denominaron las de “servicios turísticos y capacidad de establecimientos hoteleros de prestigio” y las ciudades turísticas de Almería, Huelva, Jaén y Córdoba las denominaron las de “capacidad de servicios de restauración y de establecimientos hoteleros económicos”, llegando a la mismas conclusiones con las tres técnicas Multivariantes.

Finalmente establecieron que la técnica de escalamiento multidimensional, a pesar de seguir siendo infrautilizada en muchas áreas puede ser utilizada en muchos casos como una alternativa diferente.

Corrales (2006), publica un artículo: *Escalamiento Multidimensional Para Determinar Modelos De Gestión Administrativa*, Universidad del Bío-Bío-CHILE. El objetivo de la investigación es el estudio del proceso de autoevaluación de las 7 carreras de pregrado que se han sometieron a este proceso en la Universidad del Bío-Bío, proponiendo una metodología de análisis que permite determinar cuál es la dinámica de la gestión, forma que se pueda visualizarse a través de un modelo de sistema clásico. Para poder apoyarse en dicha investigación aplicaron el análisis de escalamiento

multidimensional y así poder modelar el proceso de gestión, deduciendo su estructura asociada a un modelo de sistema realimentado.

Se consiguió formar grupos donde tuvieran alguna uniformidad como los micro criterios por ver su mayor relación y las cercanías que existan entre ellos, como el primer Grupo se consideró a los Recursos económicos ,Servicios de estudiantes ,Integridad y Organización de Administración y un segundo grupo fue considerado el de Egresado vínculos ,Currículo, Cuerpo Académico y la Metodología de la Infraestructura. Los grupos de criterios formados permitirán diseñar módulos de procesos de gestión, que analizados sobre la base del modelo de sistema, los cuales constituirán los componentes “proceso” y “realimentación” de los modelos de gestión de la actividad de autoevaluación realizados por las carreras de pregrado.

Para poder comprobar o conocer cuál de las carreras de pregrado cumplían con un criterio de similitud de gestión de procesos realizados en el procesos de autoevaluación en función de los criterios que tenían mayor semejanza respecto del proceso realizado se determinó que las carreras que realizaron proceso de gestión alguna semejanza habían sido Arquitectura, Educación Básica, Historia y Geografía y Ing. Industrial, mientras que Ing. Comercial y Enfermería presentaban alguna semejanza en su proceso de gestión.

Rangel (2012), realiza una investigación: *Evaluación de los síntomas emocionales dentro del constructo de calidad de vida en pacientes con cáncer empleando el escalamiento multidimensional*, Universidad Nacional de Colombia-COLOMBIA. El objetivo del estudio fue establecer como se relacionan estos ítems de constructo de calidad de vida en las escalas EORTC QLQC30 las cuales estaban compuestas de 30 ítems agrupados en sub-escalas; funcionamiento físico, rol emocional, cognitivo, social, estado global de la salud y calidad de vida y sub-escalas de síntomas y FACT-G compuesto de 27 ítems, distribuidos en cuatro dominios (estado físico general de la salud, ambiente familiar y social, estado emocional y capacidad de funcionamiento personal) ,aplicando este instrumento de medición de la calidad de vida en pacientes con diagnóstico de cáncer de estómago, de colon o de recto ubicados

en los servicios de hospitalización y consulta externa del Instituto Nacional de Cancerología , donde aplicaron la técnica de escalamiento multidimensional.

De los resultados que obtuvieron al aplicar el instrumento fue que en EORTC QLQ-C30 los ítems del dominio emocional mostraron una pobre correlación con los otros componentes de la escala. En FACT-G, el dominio del estado emocional mostró una estructura heterogénea y mayor correlación con los síntomas físicos.

El estudio considero que los síntomas emocionales mostraron una pobre correlación con los otros componentes del constructo. Sugiere que se debería Tener en cuenta que las escalas para medir calidad de vida en pacientes con cáncer se han desarrollado fundamentalmente en culturas anglosajonas, lo cual puede ser un factor de sesgo en el momento de aplicar el instrumento en culturas diferentes, por lo cual se valdrían ser explicarlos por diferencias culturales en la expresión emocional o cambios en el significado relacionados con la traducción.

Socha & Sigler (2012), presentan un estudio: *Using Multidimensional Scaling To Improve Functionality Of The Revised Learning Proces Questionnaire*, Western Carolina University -EE.UU. El objetivo de la investigación fue utilizar la técnica de análisis multivariante de Escalamiento Multidimensional para mejorar la funcionalidad del Cuestionario para el Proceso de Aprendizaje con la finalidad de conocer el desarrollo de una comprensión conceptual de cómo aprenden los estudiantes y lo que les motiva a participar en tareas particulares .Se consideró 329 estudiantes voluntarios en una de pública en el sureste EE.UU, consideraron primordialmente estudiar las propiedades psicométricas y la estructura factorial latente de este cuestionario para que puedan ser replicables en un contexto educativo diferente y con estudiantes de un país diferente al que originalmente utilizó para crear ,validar y poder explorar aún más la estructura la el cuestionario y mejorar la funcionalidad del diagnóstico del instrumento a través de dicha técnica.

Sánchez. & Villareal (2010), presenta un estudio: *Desarrollo de una Escala Multidimensional Breve de Ajuste Escolar*, Universidad Autónoma de Nuevo León-MEXICO. El objetivo del estudio era evaluar el grado en que el adolescente está integrado a su medio escolar, al considerar que los instrumentos hasta el momento publicados sólo contemplan el ajuste escolar como un factor dentro de un conjunto de competencias sociales o logros madurativos en los planteles públicos de educación secundaria y educación medio.(Walker & McConnell, (1995); Estévez, Musitu & Herrero, 2005) superior empleando una estructura multifactorial acorde a un concepto integral del contexto escolar en el estado de Nuevo León.

Propone una metodología (Blanco, L. & Mujica, C, 1998.) a partir de la cual se puede realizar una ordenación de objetos permitiendo la representación gráfica de medidas de proximidad entre pares de estos objetos, en la búsqueda de estructuras de conglomeración. El objetivo de definir una configuración de países latinoamericanos con respecto a características socio-demográficas y económicas. Para ello se basaron en un método de predicción aplicado por Gower en el ámbito del EM.

Se presenta un estudio (Porcar, M. & Escalante, E., 2000.) con el objetivo de estudiar los límites de la explicación interna que realizan los estudiantes de carreras universitarias de educación de la Universidad de Cuyo (Argentina) en, La perspectiva adoptada es cognitivo-afectiva y motivacional (procesos psicológicos que intervienen en la conducta de futuros docentes). Se sustenta en la necesidad de reflexionar sobre la dicotomía establecida: interno/externo entre la explicación a partir del concepto de "locus de control" Rotter (1966) y la explicación a partir de la "atribución causal", planteada por Heider (1958).

2.2.- Bases Teóricas Escalamiento Multidimensional (EM)

2.2.1.- Reseña Histórica

Esta técnica multivariante de análisis matemático es relativamente nueva, puesto que nació a mediados del siglo XX.

La cual se pueden diferenciar 4 fases o etapas en la evolución del EM:

- ✓ En esta primera fase, el nacimiento de EM, los trabajos se concentraron en lo que se conoce como “Metric MDS”. La primera proposición o estudio conocido en relación con estas técnicas se atribuye a Torgerson, W.S. que en 1952 propuso utilizar modelos Euclídeos para resolver el problema. Ya antes, en 1950 Attneave sugirió usar modelos no Euclídeos.
- ✓ La segunda etapa que se puede apreciar, destaca sobre todo por trabajos y estudios encaminados a lo que se conoce como “Non-Metric MDS”. Los investigadores más importantes que contribuyeron al desarrollo de “Non-Metric MDS” fueron, Shepard, R.N. (1962), Kruskal, J.B. (1964), Torgerson, W.S. (1962) y Guttman, L. (1968) Especial importancia a los estudios de Guttman en relación con el proceso de identificar las dimensiones y buscarles un significado.
- ✓ En la tercera etapa se sientan las bases para una variante importante de la teoría de EM, puesto que hasta ahora se había trabajado únicamente con una matriz de similitud. Ahora entran en juego las diferentes percepciones que se pueden tener frente a los mismos objetos en estudio, es lo que se conoce como “Individual Differences MDS”. Los investigadores destacados en esta etapa han sido: Tucker, G. y Messick, S.J. (1963), McGee, V.C. (1968) y Carroll, J.D. y Chang, J.J., estos últimos aportaron el primer algoritmo para trabajar con WMDS (Weighted-Multidimensional Scaling).
- ✓ En esta última etapa, los trabajos más importantes fueron dirigidos a la consolidación de estas nuevas técnicas como métodos eficaces para el análisis multivariante y estuvieron encaminadas a la unión de los distintos métodos mencionados anteriormente: Metric/Non-Metric MDS/WMDS. Los científicos más destacados, Ramsay, J.O. (1977) y de Leeuw, J. (1977).

2.2.2.-Conceptos

El EM es una herramienta matemática que usa las proximidades entre los objetos, sujetos o estímulos para producir una representación espacial de estos ítems. Las proximidades se define como un conjunto de números que expresan el valor de similaridad (similitud) o disimilaridad(desigualdad) entre los pares de objetos ,asuntos o estímulos .En comparación con otras técnicas Multivariantes de disimilaridades o distancias $D_{(n \times n)}$,con los elementos δ_{ij} y d_{ij} respectivamente.

El EM es una técnica de reducción de dimensiones porque se concentra en el problema de encontrar un conjunto de puntos de una baja dimensión que represente la “configuración” de datos en la dimensión original .la “configuración “en la dimensión original se representa por la matriz disimilaridades o distancias D.

Cada punto de la configuración corresponde a uno de los objetos. Esta configuración refleja la “estructura oculta” de los datos, y a menudo hace más fácil la comprensión y el entendimiento de los datos. Para reflejar esta estructura decimos que si la disimilitud es más grande (o la similitud más pequeña) entre dos objetos, como muestra su valor de proximidad, éstos deberían estar más apartados en el mapa de puntos.

a..- Medidas de Similitud

Existen diversas formas, pero hay tres métodos que dominan las aplicaciones delas medidas de similitud:

- Medidas de correlación,
- Medidas de asociación
- Medidas de semejanza o desemejanza (distancias).

Cada uno de los métodos representa una perspectiva particular de similitud, dependiendo tanto de sus objetivos como del tipo de datos. Tanto las medidas de distancia como la correlación exigen datos métricos, mientras que las medidas de asociación son utilizadas para datos no métricos.

b.- Coeficientes de asociación

El determinar la semejanza entre objetos o estímulos es, la matriz de los datos en una matriz cuadrada de asociación entre los objetos o estímulos. En la mayoría de los casos, la matriz de asociación es simétrica. Las matrices no-simétricas se pueden descomponer en componentes simétricos y componentes asimétricos y entonces, los componentes se pueden analizar por separado. Esto pone de relieve la importancia de elegir una medida adecuada de asociación. Esta opción se determinará por la aplicación del análisis.

Las medidas de similitud de asociación se utilizan para comparar objetos cuyas características se miden sólo en términos no métricos (nominales y ordinales). Si los datos están divididos en clases, el estadístico Chi-cuadrado es el más utilizado. Si los datos son binarios, existen una diversidad de distancias que van desde la distancia euclidiana hasta las medidas de SOKAL y Sneath, Jaccard, Lambda, Ochiai y otras y si se trata de variables ordinales o nominales, la distancia de Gower es la más conocida.

c.-Medidas de distancia

Las medidas de similitud de distancia, que representan la similitud como la proximidad de las observaciones respecto a las otras, para las variables del valor teórico del análisis del EM, son las medidas de similitud más utilizadas. El análisis de conglomerados basado en la distancia, tienen valores más parecidos para el conjunto de variables. Las medidas de distancia utilizadas para el agrupamiento pueden ser también muy diversas; entre ellas se encuentran: la distancia euclidiana, la distancia euclidiana al cuadrado, la

distancia Coseno, la distancia de Tchebychev, la de Minkowski y otras que un investigador pueda concebir para datos métricos.

d.- Medidas de proximidad y de Distancia

Una vez establecidas las variables y los objetos a clasificar el siguiente paso consiste en establecer una medida de proximidad o de distancia entre ellos que cuantifique el grado de similaridad entre cada par de objetos.

Las medidas de proximidad, similitud o semejanza miden el grado de semejanza entre dos objetos de forma que, cuanto mayor (respecto al menor) es su valor, mayor (respecto la menor) es el grado de similaridad existente entre ellos y con más (respectivamente menos) probabilidad lo métodos de clasificación tenderán a ponerlos en el mismo grupo.

Las medidas de disimilitud, de semejanza o distancia miden la distancia entre dos objetos de forma que, cuanto mayor (respecto al menor) sea su valor, más diferentes son los objetos y menor (respecto al mayor) la probabilidad de que los métodos de clasificación los pongan en el mismo grupo.

e.-Tipos de datos

- De intervalo: se trata de una matriz objetos x variables en donde todas las variables son cuantitativas, medidas en escala intervalo o razón
- Frecuencias: las variables analizadas son categóricas de forma que, por filas, tenemos objetos o categorías de objetos y, por columnas, las variables con sus diferentes categorías. En el interior de la tabla aparecen frecuencias.

- Datos binarios: se trata de una matriz objetos variables pero en la que las variables analizadas son binarias de forma que 0 indica la ausencia de una característica y 1 su presencia.

2.3.3.- Medidas de proximidad

f.-Medidas para Variables Cuantitativas

G.- Coeficiente de Congruencia

$$C_{rs} = \frac{\sum_{j=1}^p X_{rj} X_{sj}}{\sqrt{\sum_{j=1}^p X_{rj}^2} \sqrt{\sum_{j=1}^p X_{sj}^2}}$$

Que es el coseno del ángulo que forman los vectores $(x_{r1}, \dots, x_{rp})'$ y $(x_{s1}, \dots, x_{sp})'$.

h.-Coeficiente de Correlación

$$r_{rs} = \frac{\sum_{j=1}^p (X_{rj} - \bar{X}_r)(X_{sj} - \bar{X}_s)}{\sqrt{\sum_{j=1}^p (X_{rj} - \bar{X}_r)^2} \sqrt{\sum_{j=1}^p (X_{sj} - \bar{X}_s)^2}}$$

Dónde:

$$\bar{X}_r = \frac{\sum_{j=1}^p X_{rj}}{p} \quad \text{y} \quad \bar{X}_s = \frac{\sum_{j=1}^p X_{sj}}{p}$$

Si los objetos r y s son variables, r_{rs} mide el grado de asociación lineal existente entre ambas.

Estas dos medidas se utilizan, preferentemente para clasificar variables siendo, en este caso, invariantes por cambios de escala y, en el caso del coeficiente de correlación, invariante por

cambio de origen. Por esta razón es más conveniente utilizar el coeficiente de congruencia con variables tipo razón en las cuales el origen está claramente definido.

i.-Medidas para Datos Binarios

En este caso se construye una tabla de contingencia, para cada par de objetos r y s , de la forma:

Tabla N° 1: Tabla de Contingencia para objetos r y s		
Objeto $s \setminus$ Objeto r	0	1
0	a	b
1	c	d

Fuente: Hair, Anderson, Tatham y Black, 2000

Donde a = número de variables en las que los objetos r y s toman el valor 0, etc. y $p = a+b+c+d$. Utilizando dichas tablas algunas de las medidas de semejanza más utilizadas son:

$$\text{Coeficiente de Jaccard: } \frac{d}{b + c + d}$$

$$\text{Coeficiente de Acuerdo Simple: } \frac{a + d}{p}$$

Tabla N°2: Medidas de Similaridad Datos Binarios

SIMILARIDAD	$D = 1 - S$ Métrica, etc.	$D = 1 - S$ Euclídeana	$D = \sqrt{1 - S}$ Métrica	$D = \sqrt{1 - S}$ Euclídeana
$s_1 = \frac{a+d}{a+b+c+d}$ <p><i>Emparejamiento Simple</i></p>	Métrica	NO	SI	SI
$s_2 = \frac{a+d}{a+2b+2c+d}$ <p><i>Roger & Tanimoto</i></p>	Métrica	NO	SI	SI
$s_3 = \frac{2a+2d}{2a+b+c+2d}$	Semimétrica	NO	SI	NO
$s_4 = \frac{a+d}{b+c}$	No métrica	NO	NO	NO
$s_5 = \frac{1}{4} \left[\frac{a}{a+b} + \frac{a}{a+c} + \frac{d}{b+d} + \frac{d}{c+d} \right]$	Semimétrica	NO	NO	NO
$s_6 = \frac{a}{\sqrt{(a+b)(a+c)}} \frac{d}{\sqrt{(b+d)(c+d)}}$	Semimétrica	NO	SI	SI
$s_7 = \frac{a}{a+b+c}$ <p><i>Jaccard</i></p>	Métrica	NO	SI	SI
$s_8 = \frac{2a}{2a+b+c}$	Semimétrica	NO	SI	SI
$s_9 = \frac{3a}{3a+b+c}$	Semimétrica	NO	NO	NO
$s_{10} = \frac{a}{a+2b+2c}$	Métrica	NO	SI	SI

$s_{11} = \frac{a}{a+b+c+d}$ <small>Rosset & Rao</small>	Métrica	NO	SI	SI
$s_{12} = \frac{a}{b+c}$	No Métrica	NO	NO	NO
$s_{13} = \frac{1}{2} \left[\frac{a}{a+b} + \frac{a}{a+c} \right]$	Semimétrica	NO	NO	NO
$s_{14} = \frac{a}{\sqrt{(a+b)(a+c)}}$ <small>Ochiai</small>	Semimétrica	NO	SI	SI
$s_{15} = \sum w_j s_j / \sum w_j$ <small>Gowert</small>	Métrica	NO	SI	Como* (S1)
$s_{16} = \sum w_j s_j / \sum w_j$ <small>Estabrook & Rogers</small>	Métrica	NO	SI	Como* (S1)
$s_{17} = \frac{2W}{A+B}$ <small>Stenhaus</small>	Semimétrica	NO	Como* (S1)	Como* (S1)
$s_{18} = \frac{1}{2} \left[\frac{W}{A} + \frac{W}{B} \right]$ <small>Kuczyński</small>	Semimétrica	NO	No* (S13)	No* (S13)
$s_{19} = \sum w_j s_s / \sum w_j$ <small>Gowert</small>	Métrica	NO	SI	Como* (S7)
$s_{20} = \sum w_j s_s / \sum w_j$ <small>Legendre & Chodorowski</small>	Métrica	NO	SI	Como* (S7)
$s_{21} = 1 - \chi^2$	Métrica	SI	SI	SI

Fuente: Gower & Legendre, 1986

2.3.5.- Modelo General EM

Existen dos modelos básicos de EM que son: el modelo de escalamiento métrico y el modelo de escalamiento no métrico. En el primero de ellos consideramos que los datos están medidos en escala de razón o en escala de intervalo y en el segundo consideramos que los datos están medidos en escala ordinal. No se ha desarrollado todavía ningún modelo para datos en escala nominal.

j.- Modelo de EM Métrico.

Supongamos que tenemos n con las disimilaridades $\{\delta_n\}$, EM métrico intenta encontrar un conjunto de puntos en un espacio donde cada punto representa uno de los objetos y las distancias entre los puntos $\{d_n\}$ es tal que:

$$d_{rs} \approx f(\delta_n)$$

Donde la f es una función paramétrica monótona continua .la función f puede ser una función identidad o una función que intenta transformar las disimilaridades en forma de distancia.

l.- Escalamiento Clásico

El escalamiento clásico se originó en los años treinta cuando (Young & Huseholder, 1983), mostro como abordando con una matriz de distancia entre los puntos en un espacio euclidiano, pueden encontrarse coordenadas para los puntos tal que las distancias se conservan.

En realidad, es bastante subjetivo el hecho de elegir una medida de similitud ya que depende de las escalas de medida. Se pueden agrupar observaciones según la similitud expresada en términos de una distancia. Si se agrupan variables, es habitual utilizar como medida de similitud los coeficientes de correlación en valor absoluto. Para variables categóricas existen también

criterios basados en la posesión o no de los atributos (tablas de presencia-ausencia).

Dados dos vectores X_i, X_j pertenecientes a r_k , diremos que hemos establecido una distancia entre ellos si definimos una función d con las propiedades siguientes:

1. $d: n_k \times n_k \rightarrow r^+$, es decir $d(X_i, X_j) \geq 0$;
2. $d(X_i, X_j) = 0$, la distancia entre un elemento y sí mismo es cero.
3. $d(X_i, X_j) = d(X_j, X_i)$, la distancia es simétrica
4. $d(X_i, X_j) \leq d(X_i, X_p) + d(X_j, X_p)$ la distancia verifica la propiedad triangular.

Estas propiedades generalizan la noción intuitiva de distancia euclídea entre dos puntos. Si sólo se cumplen las dos primeras condiciones, diremos que $\delta_{(i; j)}$ es una disimilaridad de modo general, podemos decir que el EM toma como entrada una matriz de proximidades, $\delta \in \mu_{n \times n}$, donde n es el número de estímulos. Cada elemento δ_{ij} de δ representa la proximidad entre el estímulo i y el estímulo j .

$$\Delta = \begin{pmatrix} \delta_{11} & \dots & \delta_{1n} \\ \vdots & \ddots & \vdots \\ \delta_{n1} & \dots & \delta_{nn} \end{pmatrix}$$

$$\delta_{ij} = \delta_{ji} = \delta(i, j) \geq 0$$

Diremos que $\Delta = (\delta_{ij})$ es una matriz de distancias euclídeas si existen n puntos $x_1, \dots, x_n \in \mathbb{R}^p$; siendo:

$$X'_i = (x_{i1}, \dots, x_{ip}) \quad i = 1, \dots, n$$

$$\text{Tales que: } \delta_{ij}^2 = \sum_{\alpha=1}^p (x_{i\alpha} - x_{j\alpha})^2 = (X_i - X_j)'(X_i - X_j)$$

A partir de esta matriz de proximidades el EM nos proporciona como salida una matriz $X \in \mathbb{R}^{n \times m}$, donde n , es el número de estímulos, y m es el número de dimensiones. Cada valor x_{ij} representa la coordenada del estímulo i en la dimensión.

$$X = \begin{pmatrix} X_{11} & \dots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{nm} \end{pmatrix}$$

El objetivo del escalado multidimensional es encontrar la X más adecuada a partir de la matriz de distancias Δ donde p puede ser un valor entre 1 e infinito. A partir de estas distancias podemos obtener una matriz de distancias que denominamos $D \in \mathbb{R}^{n \times n}$:

$$D = \begin{pmatrix} d_{11} & \dots & d_{1n} \\ \vdots & \ddots & \vdots \\ d_{n1} & \dots & d_{nn} \end{pmatrix}$$

La solución proporcionada por el EM debe ser de tal modo que haya la máxima correspondencia entre la matriz de proximidades inicial δ y la matriz de distancias obtenidas D . para que exista la máxima correspondencia EM proporciona varias medidas, que veremos más adelante, y que nos informan sobre la bondad del modelo.

m.- Distancia Euclídea.

Dados dos objetos I_1 y I_2 medidos según dos variables x_1 y x_2 , la distancia euclídea entre ambos es:

$$d_{I_1 I_2} = \sqrt{(x_{11} - x_{21})^2 + (x_{12} - x_{22})^2}$$

Con más dimensiones (o variables que se miden) es equivalente a:

$$d_{I_1 I_2} = \sqrt{\sum_{k=1}^p (x_{1k} - x_{2k})^2}$$

En notación vectorial se expresa como:

$$d_{I_i I_j}^2 = (x_i - x_j)' + (x_i - x_j)$$

Si se consideran n objetos para $i, j \in \{1, \dots, n\}$ la distancia total es:

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^p (x_{ik} - x_{jk})^2)^{1/2}$$

n.- Distancia de Minkowski.

$$d_{I_i I_j} = \left[\sum_k |x_{ik} - x_{jk}|^m \right]^{1/m}, \text{ donde } m \in \mathbb{N}.$$

Si $m=1$, se tiene la distancia en valor absoluto y si $m=2$, la euclídea.

ñ.- Distancia de Mahalanobis

Se define como:
$$d_{I_i I_j}^2 = (x_i - x_j)' W^{-1} (x_i - x_j)$$

Donde w es la matriz de covarianzas entre las variables. De este modo, las variables se ponderan según el grado de relación que exista entre ellas, es decir, si están más o menos correlacionadas. si la correlación es nula y las variables están estandarizadas, se obtiene la distancia euclídea.

2.3.6.- Recuperación de Coordenadas

El método para encontrar las coordenadas euclidianas originales de las distancias euclidianas derivadas, fue propuesto (Schoenberg, 1935, Young &Householder ,1938).

Sean las coordenadas de los n puntos en un espacio euclidiano de p-dimensiones, es dado por $X_r (r = 1, \dots, n)$ donde $X_r = (X_{r1} \dots X_{rp})^T$, Entonces la distancia euclidiana entre el r-esimo y s-esimo punto es dado por:

$$d_{rs}^2 = (X_r - X_s)^T (X_r - X_s)$$

Sea la matriz de producto interno b, donde

$$[B]_{rs} = b_{rs} = x_r^T x_s$$

De la distancia al cuadrado (d_{rs}^2), la matriz de producto interno b es calculada, donde b es de coordenadas desconocidas.

Para encontrar b

$$\sum_{r=1}^n x_{ri} = 0 (i = 1, \dots, n)$$

Para encontrar b, tenemos

$$d_{rs}^2 = x_r^T x_r + x_s^T x_s - 2x_r^T x_s$$

Dónde:

$$\frac{1}{n} \sum_{r=1}^n d_{rs}^2 = \frac{1}{n} \sum_{r=1}^n x_r^T x_r + x_s^T x_s$$

$$\frac{1}{n} \sum_{r=1}^n d_{rs}^2 = x_r^T x_r + \frac{1}{n} \sum_{s=1}^n x_s^T x_s$$

$$\frac{1}{n} \sum_{r=1}^n \sum_{s=1}^n d_{rs}^2 = \frac{2}{n} \sum_{r=1}^n x_r^T x_r$$

Sustituyendo dentro de d_{rs}^2 , tenemos

$$d_{rs}^2 = x_r^T x_s$$

$$= -\frac{1}{2} \left(d_{rs}^2 - \frac{1}{n} \sum_{r=1}^n d_{rs}^2 - \frac{1}{n} \sum_{s=1}^n d_{rs}^2 + \frac{1}{n^2} \sum_{r=1}^n \sum_{s=1}^n d_{rs}^2 \right)$$

$$= a_{rs} - a_r - a_s + a$$

Definimos la matriz a como: $[A]_{rs} = a_{rs}$

$$\text{Donde : } a_{rs} = -\frac{1}{2} d_{rs}^2$$

$$a = n^{-1} \sum_s a_{rs}, \quad \text{es la distancia cuadrática media por fila,}$$

$a_s = n^{-1} \sum_s a_{rs}$, es la distancia cuadrática media por columna,

$a = n^{-2} \sum_r \sum_s a_{rs}$, es la distancia cuadrática media de la matriz.

La matriz de producto interno ahora es: $B = HAH$

Donde H es la matriz de centricidad: $H = 1 - n^{-1} 11^T$

la matriz producto interno puede ser escrita como: $B = XX^T$

$X = [x_1, \dots, x_n]^T$, es una matriz de coordenadas $n \times p$, b es simétrica, semi definida positiva.

(b), es dado por:

$$r(B) = r(XX^T) = r(X) = p$$

Donde p son autovalores no negativos y $(n-p)$ autovalores iguales a cero. La matriz b , se puede escribirse en término de la descomposición espectral.

$B = V \Lambda V^T$ $V = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, que es la diagonal de la matriz de autovalores $\{\lambda_i\}$ de b , los autovalores de b son de tal que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ y $V = [v_1, v_2, \dots, v_n]$ es su correspondiente matriz de

autovectores, estandarizando tal que $v_i^T v_i = 1$.

Debido a los $n-p$ autovalores iguales a cero, b puede ser escrito como:

$$B = V_1 \Lambda_1 V_1^T$$

Dónde: $\Lambda_1 = \text{diag}(\lambda_1, \dots, \lambda_p)$, $V = v_1, v_2, \dots, v_p$

Como $B = XX^T$, la matriz de coordenadas x es dado por: $X = V_1 \Lambda_1^{1/2}$

Donde $\Lambda_1^{1/2} = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_p^{1/2})$, de esta manera se han recuperado las coordenadas de las distancias entre los puntos.

Todo modelo de escalamiento parte de la idea de que las distancias son una función de las proximidades, es decir $d_{rs} \approx f(\delta_{rs})$. En el modelo de escalamiento métrico partimos del supuesto de que la relación entre las proximidades y las distancias es de tipo lineal:

$d_{rs} \approx a + b(\delta_{rs})$, en el cual a partir de una matriz de, d , se puede obtener una matriz b de productos escalares entre vectores. el procedimiento consiste en transformar la matriz de proximidades Δ en una matriz de distancias d , de tal forma que verifique los tres axiomas de la distancia euclidiana:

No negatividad $d_{rs} \geq 0 = d_{rr}$

Simetría $d_{rs} = d_{sr}$

Desigualdad triangular $d_{rs} \leq d_{ri} + d_{is}$

Los dos primeros axiomas son fáciles de cumplir, pero el tercer axioma no se cumple siempre. Este problema se conoce con el nombre de “estimación de la constante aditiva”. Torgerson encontró la solución a este problema, estimando el valor mínimo de c que verifica la desigualdad triangular de la siguiente

forma: $c_{\min} = \max_{rsi} \{\delta_{rs} - \delta_{ri} - \delta_{is}\}$

De esta forma las distancias se obtienen sumando a las proximidades la constante c , es decir. $d_{rs} = \delta_{rs} + c$

Una vez obtenida la matriz d es necesario transformarla en una matriz b de productos escalares entre vectores mediante la siguiente transformación:

$$b_{rs} = -\frac{1}{2}(a_{rs}^2 - a_{r.}^2 - a_{.s}^2 + a_{..}^2)$$

Luego, lo único que queda es transformar la matriz b en una matriz x tal que $b=xx^t$, teniendo x la matriz que nos da las coordenadas de cada uno de los n estímulos en cada una de las p dimensiones. Cualquier método de factorización permite transformar b en x . En resumen el procedimiento consiste en transformar:

Proximidades \rightarrow Distancias \rightarrow Productos \rightarrow Escalares \rightarrow Coordenadas

2.3.7.- EM No Métrico.

A diferencia del EM métrico, el modelo de EM no métrico no presupone una relación lineal entre las proximidades y las distancias, sino que establece una relación monótona creciente entre ambas, es decir, si $\delta_{ij} < \delta_{kl} \Rightarrow d_{ij} < d_{kl}$. Su desarrollo se debe a Shepard (1962) quién demostró que es posible obtener soluciones métricas asumiendo únicamente una relación ordinal entre proximidades y distancias. Posteriormente Kruskal (1964) mejoró el modelo. El procedimiento se basa en los siguientes apartados:

- Transformación de la matriz de proximidades en una matriz de rangos, desde 1 hasta $(n(n-1))/2$.
- Obtención de una matriz $X \in M_{n \times m}$ de coordenadas aleatorias, que nos da la distancia entre los estímulos.
- Comparación de las proximidades con las distancias, obteniéndose las disparidades (d'_{ij}).
- Definición del Stress.
- Minimización del Stress, Tanto para el modelo métrico como para el modelo no métrico es necesario obtener un coeficiente que nos informe sobre la bondad

del modelo. Sabemos que las distancias son una función de las proximidades, es decir:

$$f: \delta_{ij}(x) \rightarrow d_{ij}(x)$$

De esta forma se tiene que $d_{ij}=f(\delta_{ij})$. Esto no deja ningún margen de error, sin embargo, en las proximidades empíricas es difícil que se dé la igualdad, con lo que generalmente ocurre que $d_{ij} \approx f(\delta_{ij})$. A las transformaciones de las proximidades por f se le denomina disparidades. A partir de aquí podemos definir el error cuadrático como:

$$e_{ij}^2 = (f(\delta_{ij}) - d_{ij})^2$$

Como medida que nos informa de la bondad del modelo podemos utilizar el Stress que Kruskal definió como:

$$Stress = \sqrt{\frac{\sum_{i,j} (f(\delta_{ij}) - d_{ij})^2}{\sum_{i,j} d_{ij}^2}}$$

Mientras mayor sea la diferencia entre las disparidades y las distancias, es decir, entre $f(\delta_{ij})$ y d_{ij} , mayor será el Stress y por tanto peor será el modelo. Por tanto, el Stress no es propiamente una medida de la bondad del ajuste, sino una medida de la no bondad del ajuste. Su valor mínimo es 0, mientras que su límite superior para n estímulos es $1 - (2/n)$.

Kruskal (1964) sugiere las siguientes interpretaciones del Stress:

- 0.15 - 0.20 → Pobre
- 0.10 - 0.15 → Aceptable
- 0.05 - 0.10 → Bueno
- 0.01 - 0.05 → Aceptable
- 0.0 - 0.01 → Excelente

También se suele utilizar una variante del Stress que se denomina S-Stress, definida

Como:

$$S - Stress = \sqrt{\frac{\sum_{i,j} (f(d_{ij})^2 - d_{ij}^2)^2}{\sum_{i,j} (d_{ij}^2)^2}}$$

Otra medida que se suele utilizar es el coeficiente de correlación al cuadrado (RSQ), que nos informa de la proporción de variabilidad de los datos de partida que es explicada por el modelo. Los valores que puede tomar oscilan entre 0 y 1, al ser un coeficiente de correlación al cuadrado. Valores cercanos a 1 indican que el modelo es bueno y valores cercanos a 0 indican que el modelo es malo. Su expresión es:

$$RSQ = \sqrt{\frac{[\sum_i \sum_j (d_{ij} - d_{..})(f(d_{ij}) - f(d_{..}))]^2}{[\sum_i \sum_j (d_{ij} - d_{..})^2][\sum_i \sum_j (f(d_{ij}) - f(d_{..}))^2]}}$$

La mayoría de los paquetes estadísticos tienen implementados tanto los algoritmos para obtener soluciones con MDS así como las medidas para determinar si el modelo es adecuado o no¹. En la actualidad todo los

algoritmos implementados en los paquetes estadísticos son reiterativos, de forma que se alcance la mejor solución posible..

2.3.8.- Número de Dimensiones

Los autovalores $\{\lambda_i\}$ indican cuantas dimensiones se requieren para representar las similaridades $\{\delta_{is}\}$, si b es entonces semi-definido positivo entonces el número de autovalores diferentes de cero, da el número de dimensiones requeridos, si b no es semi-definido positivo, entonces el número de autovalores positivo es el número propio de dimensiones. Este son las dimensiones máximas del espacio requeridas. Sin embargo, en la práctica, el número de dimensiones necesarias pueden ser escogidas, buscando la menor dimensión (el cual es normalmente 2 o 3).

La suma de las distancias al cuadrado entre puntos en el espacio completo es:

$$\frac{1}{2} \sum_{r=1}^n \sum_{s=1}^n d_{rs}^2 = n \sum_{r=1}^n x_r^T x_r = n \left(\text{tr}(B) \right) = n \sum_{i=1}^{n-1} \lambda_i$$

Una medida de la proporción de variabilidad explicada usando sólo p dimensiones es:

$$\frac{\sum_{i=1}^p \lambda_i}{\sum_{i=1}^{n-1} \lambda_i}$$

Si b , no es semi definida positiva esta medida de variabilidad es modifica por:

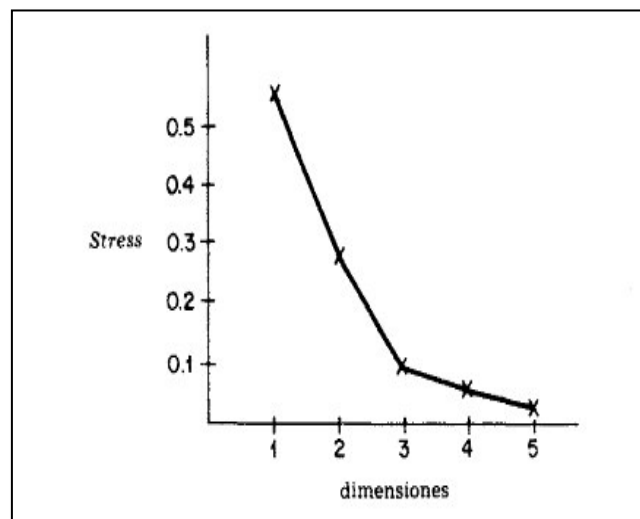
$$\frac{\sum_{i=1}^p \lambda_i}{\sum_{i=1}^{n-1} |\lambda_i|}$$

Si B , no es semi definida positiva esta medida de variabilidad es modifica por:

$$\frac{\sum_{i=1}^p \lambda_i}{\sum_{i=1}^{n-1} |\lambda_i|} \quad \text{o} \quad \frac{\sum_{i=1}^p \lambda_i}{\sum_{i=1}^p \lambda_i} \quad (\text{autovalores} - \text{positivos})$$

Para poder determinar del número de dimensiones que uno quiere considerar, no hay una regla predeterminada, muchas veces el investigador por propia experiencia ó conocimiento a priori puede determinar con el número de dimensiones que quiere trabajar ó hacer un análisis exploratorio con que puede ser mediante el gráfico de la Ladera donde se encuentre el punto de inflexión nos va indicar el número de dimensiones, también se puede determinar verificando el valor Stress, para diferentes las diferentes número de dimensiones observar su punto de inflexión en el número de dimensiones.

Figura N°1: Valor del Stress según el número de Dimensiones.



Fuente : Hair & Anderson & Tathan & Black

Capítulo III:

Metodología

3.1.- Tipo y Diseño de Investigación

La presente investigación es de tipo Aplicada -descriptivo y diseño No Experimental.

3.2.- Unidad de Análisis

La unidad de análisis se encuentra conformada por una Institución Educativa del nivel Secundaria del Perú.

3.3.- Población de Estudio

Conformada por toda las Instituciones Educativas del nivel Secundaria del Perú.

3.4.-Tamaño de Muestra

La Organización para la Cooperación y el Desarrollo Económico (OCDE), a través del Programa Internacional de Evaluación de Estudiantes (PISA), consideró seleccionar a 240 Instituciones educativas del nivel secundaria del Perú.

3.5.- Selección de Muestra

La Coordinación del Programa Internacional de Evaluación de Estudiantes (PISA) – Perú, envía el Marco Muestral de la IIEE a la Organización para la

Cooperación y el Desarrollo Económico (OCDE), quien realizan la selección a través de estratos explícitos por gestión y área de la IE: Estatal-Urbana, Estatal-Rural, No estatal-Urbana, No estatal-Rural.

3.6.- Técnica de Recolección de Datos

La técnica de recolección de datos se dio a través de la información de fuentes Secundarias.

3.7.- Mapas Perceptuales

Los mapas perceptuales son un panorama más visual del lugar que ocupa una empresa o producto con respecto a los competidores y de la percepción que tiene el cliente de la misma. (Serralvo & adeu, 2005, p. 3). Las posiciones de los objetos (marcas, productos, tipos de consumidor), en las dimensiones relevantes que describen el mercado en el cual participan, se representan gráficamente en un mapa perceptual.

La interpretación de los ejes (dimensiones) y de las distancias entre los puntos del mapa perceptual depende de la técnica de posicionamiento a partir de la cual fueron elaborados. (Serralvo y Tadeu, 2005, p. 3)

Los mapas perceptuales son de gran utilidad en el Marketing ya que nos ayudan a tener una idea más clara de cómo están ubicadas o consideradas las distintas marcas en la mente del consumidor, en otras palabras, nos otorgan un mayor conocimiento del posible posicionamiento que pueden tener. Por ende, permite a las empresas realizar un seguimiento y realizar campañas que apoyen ese posicionamiento o reposicionaren el caso de que posea algo negativo.

3.10.-El Programa Internacional para la Evaluación de Estudiantes (PISA) de la Organización para la Cooperación y el Desarrollo Económico (OCDE)

El Programa Internacional para la Evaluación de Estudiantes (PISA) es una iniciativa desarrollada por la OCDE principalmente orientada a los países que forman parte de dicha organización. Su propósito fundamental es conocer el nivel de competencias básicas de la población de quince años, es decir, de aquella población que en dichos países se encuentra próxima a terminar la educación obligatoria e iniciar estudios post secundarios o incorporarse a la vida laboral. Sin embargo, desde la primera aplicación de PISA en el año 2000, varios países que no eran miembros de la OCDE –entre ellos el Perú– expresaron su interés en participar. A partir de ello, fueron admitidos bajo la categoría de “países socios en PISA”. Ello ha hecho que, actualmente, PISA incluya a un número significativo de estos países, aun cuando el número y características de los que participan en cada ciclo varíe. En el Programa Internacional de Evaluación de Estudiantes (PISA) 2012, participaron 65 países.

El objetivo de Programa Internacional de Evaluación de Estudiantes (PISA) es conocer en qué medida los estudiantes de quince años de edad son capaces de utilizar los conocimientos y habilidades que han desarrollado, y que están relacionadas a las áreas de Lectura, Matemática y Ciencia, para hacer frente a las situaciones y desafíos que les plantea la sociedad actual con el fin de que puedan participar de manera pertinente en ella. Estas competencias han sido seleccionadas por un grupo de expertos, para lo cual se ha tomado como referencia los estudios y tendencias mundiales sobre el tipo de habilidades que se espera sean requeridos en las siguientes décadas.

Puesto que dichos conocimientos y habilidades son el resultado de diversas circunstancias personales y familiares como sociales y culturales, dentro y fuera del contexto escolar, PISA se aproxima a dichas situaciones recogiendo información a través de cuestionarios aplicados directamente a los actores escolares; asimismo, utiliza otras fuentes secundarias disponibles. Con este conjunto de datos, PISA busca aportar información relevante sobre los sistemas educativos para la formulación y discusión de políticas educativas.

La aplicación de Programa Internacional de Evaluación de Estudiantes (PISA) tiene una periodicidad regular y busca ofrecer resultados comparables a lo largo del tiempo. Está organizado en tres ciclos evaluativos de tres años cada uno. En cada ciclo, se estudia a profundidad una de las competencias evaluadas, que se traduce en la aplicación de un mayor número de preguntas o ítems de la competencia privilegiada en relación con las otras dos. Ello permite reportar los resultados de esta no solo en términos más exhaustivos, sino también en términos de sub-escalas (habilidades o áreas específicas dentro de la competencia mayor). Este diseño permite tener indicadores sobre la evolución del desempeño de cada país en las competencias evaluadas, y para el conjunto de países participantes.

Es importante tener en cuenta que el Programa Internacional de Evaluación de Estudiantes (PISA) fue originalmente concebido a finales de los años 90 en el marco de crecientes preocupaciones y el surgimiento de importante evidencia acerca de que, si bien la mayor parte de los países de la OCDE garantizaban la universalidad de la educación obligatoria, estos niveles de acceso no se traducían necesariamente en aprendizajes claves para el funcionamiento de las personas en su vida adulta.

De esta manera, Programa Internacional de Evaluación de Estudiantes (PISA) fue pensado como un estudio que daría cuenta de las competencias básicas de la población de quince años en su conjunto. Debido a que en la mayor parte de los países de la OCDE esa población se encontraba escolarizada, era posible pensar en una estrategia de investigación que se aproximara a ellos en las escuelas secundarias.

Sin embargo, en algunos países de la OCDE (como México) y otros que se sumaron al estudio, esto no era así. Por lo mismo, se redefinió parcialmente el objetivo del estudio, de modo que PISA es hoy concebida como una iniciativa que da cuenta de las competencias básicas de las personas de 15 años matriculadas en algún programa educativo de séptimo grado o superior.³ Es decir, Programa Internacional de Evaluación de Estudiantes (PISA) es, principalmente, un estudio de los estudiantes de quince años matriculados en la educación secundaria. (http://umc.minedu.gob.pe/wp-content/uploads/2013/12/reporte_pisa_2012.pdf, 2013)

El Perú, se someta a una evaluación Internacional como Programa Internacional de Evaluación de Estudiantes (PISA), al ser un país invitado para participar en la evaluación puede participar o no. Sin embargo competimos con los países miembros de la Organización de Cooperación para el Desarrollo Económico es decir con los países desarrollados donde invierten mayor presupuesto en educación. Cabe indicar que dicha evaluación no es un concurso inter-colegios o inter-países, aunque la prensa le dé a conocer en ese tono, es la apreciación sobre la realidad educativa en cada país, países con diversas realidades políticas, económicas, sociales y culturales.

Se debe buscar estrategias para salir adelante y es un reto para los maestros y la sociedad, salirse es rendirse y no aceptar el resultado, Programa Internacional de Evaluación de Estudiantes (PISA) no mide conocimientos, sino cuanto estamos preparados para hacer frente a la vida. (<http://www.perueduca.pe/foro>)

Capítulo IV:

Resultados y Discusión

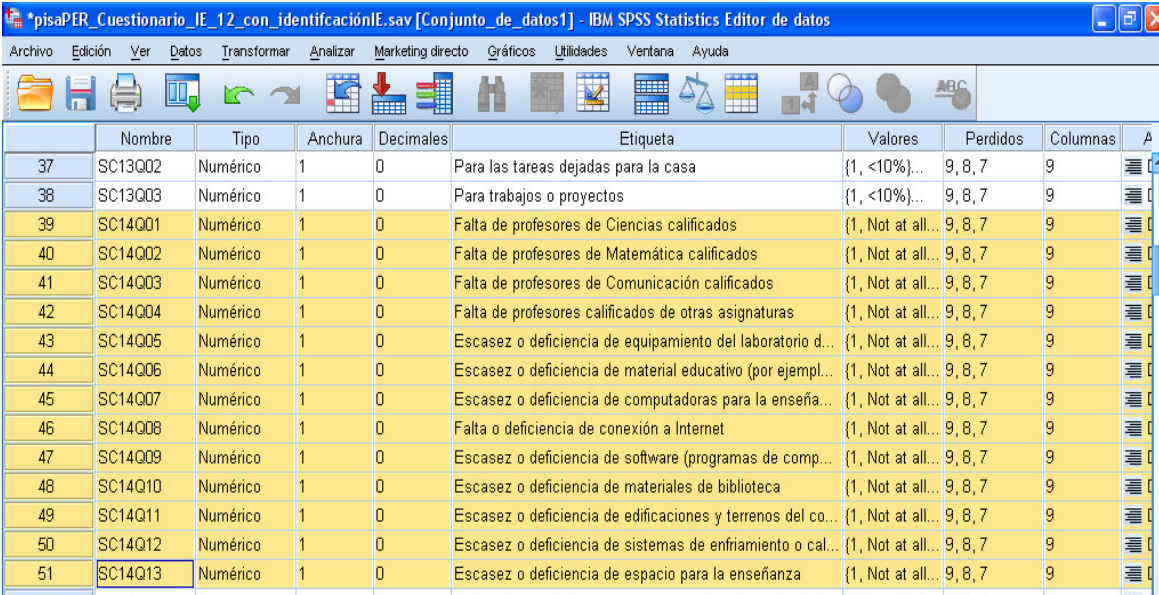
Aplicación Del EM No Métrico

En lo que sigue, se analizarán los pasos a seguir para llevar a cabo el EM No Métrico, en la base de datos PISA del año 2012 a través del software SPSS.

4.1.- Análisis, interpretación y discusión de resultados

4.1.1.-Selección de variables de la Base de datos PISA

Figura N°2: Selección de Variables de la Base de Datos PISA.



	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	A
37	SC13Q02	Númerico	1	0	Para las tareas dejadas para la casa	{1, <10%}...	9, 8, 7	9	
38	SC13Q03	Númerico	1	0	Para trabajos o proyectos	{1, <10%}...	9, 8, 7	9	
39	SC14Q01	Númerico	1	0	Falta de profesores de Ciencias calificados	{1, Not at all...	9, 8, 7	9	
40	SC14Q02	Númerico	1	0	Falta de profesores de Matemática calificados	{1, Not at all...	9, 8, 7	9	
41	SC14Q03	Númerico	1	0	Falta de profesores de Comunicación calificados	{1, Not at all...	9, 8, 7	9	
42	SC14Q04	Númerico	1	0	Falta de profesores calificados de otras asignaturas	{1, Not at all...	9, 8, 7	9	
43	SC14Q05	Númerico	1	0	Escasez o deficiencia de equipamiento del laboratorio d...	{1, Not at all...	9, 8, 7	9	
44	SC14Q06	Númerico	1	0	Escasez o deficiencia de material educativo (por ejempl...	{1, Not at all...	9, 8, 7	9	
45	SC14Q07	Númerico	1	0	Escasez o deficiencia de computadoras para la enseña...	{1, Not at all...	9, 8, 7	9	
46	SC14Q08	Númerico	1	0	Falta o deficiencia de conexión a Internet	{1, Not at all...	9, 8, 7	9	
47	SC14Q09	Númerico	1	0	Escasez o deficiencia de software (programas de comp...	{1, Not at all...	9, 8, 7	9	
48	SC14Q10	Númerico	1	0	Escasez o deficiencia de materiales de biblioteca	{1, Not at all...	9, 8, 7	9	
49	SC14Q11	Númerico	1	0	Escasez o deficiencia de edificaciones y terrenos del co...	{1, Not at all...	9, 8, 7	9	
50	SC14Q12	Númerico	1	0	Escasez o deficiencia de sistemas de enfriamiento o cal...	{1, Not at all...	9, 8, 7	9	
51	SC14Q13	Númerico	1	0	Escasez o deficiencia de espacio para la enseñanza	{1, Not at all...	9, 8, 7	9	

Fuente: OCDE-PISA.

Figura N°3: Selección de Variables de la Base de Datos PISA.

39	Falta de profesores de Ciencias calificados	{1, Not at all...	9, 8, 7	9	Derecha	Ordinal	Entrada
40	Falta de profesores de Matemática calificados	{1, Not at all...	9, 8, 7	9	Derecha	Ordinal	Entrada
41	Falta de profesores de Comunicación calificados	{1, Not at all...	9, 8, 7	9	Derecha	Ordinal	Entrada
42	Falta de profesores calificados de otras asignaturas	{1, Not at all...	9, 8, 7	9	Derecha	Ordinal	Entrada
43	Escasez o deficiencia de equipamiento del laboratorio d...	{1, Not at all...	9, 8, 7	9	Derecha	Ordinal	Entrada
44	Escasez o deficiencia de material educativo (por ejempl...	{1, Not at all...	9, 8, 7	9	Derecha	Ordinal	Entrada
45	Escasez o deficiencia de computadoras para la enseña...	{1, Not at all...	9, 8, 7	9	Derecha	Ordinal	Entrada
46	Falta o deficiencia de conexión a Internet	{1, Not at all...	9, 8, 7	9	Derecha	Ordinal	Entrada
47	Escasez o deficiencia de software (programas de comp...	{1, Not at all...	9, 8, 7	9	Derecha	Ordinal	Entrada
48	Escasez o deficiencia de materiales de biblioteca	{1, Not at all...	9, 8, 7	9	Derecha	Ordinal	Entrada
49	Escasez o deficiencia de edificaciones y terrenos del co...	{1, Not at all...	9, 8, 7	9	Derecha	Ordinal	Entrada
50	Escasez o deficiencia de sistemas de enfriamiento o cal...	{1, Not at all...	9, 8, 7	9	Derecha	Ordinal	Entrada
51	Escasez o deficiencia de espacio para la enseñanza	{1, Not at all...	9, 8, 7	9	Derecha	Ordinal	Entrada

Fuente: OCDE-PISA.

Tabla N° 3.-Variables Seleccionadas para el Análisis EM No Métrico

SC13Q02	Para las tareas dejadas para la casa
SC13Q03	Para trabajos o proyectos
SC14Q01	Falta de profesores de Ciencias calificados
SC14Q02	Falta de profesores de Matemática calificados
SC14Q03	Falta de profesores de Comunicación calificados
SC14Q04	Falta de profesores calificados de otras asignaturas
SC14Q05	Escasez o deficiencia de equipamiento del laboratorio de Ciencias
SC14Q06	Escasez o deficiencia de material educativo (por ejemplo, libros de texto)
SC14Q07	Escasez o deficiencia de computadoras para la enseñanza
SC14Q08	Falta o deficiencia de conexión a Internet
SC14Q09	Escasez o deficiencia de software (programas de computadora) para la enseñanza
SC14Q10	Escasez o deficiencia de materiales de biblioteca
SC14Q11	Escasez o deficiencia de edificaciones y terrenos del colegio
SC14Q12	Escasez o deficiencia de sistemas de enfriamiento o calentamiento y sistemas de electrificación
SC14Q13	Escasez o deficiencia de espacio para la enseñanza
SC22Q02	Estudiantes que no asisten a algunas de las clases durante la jornada escolar
SC22Q03	La dificultad de aprendizaje - Estudiantes llegando tarde
SC22Q04	Estudiantes que llegan tarde al colegio
SC22Q05	Estudiantes que no asisten a eventos escolares obligatorios (por ejemplo, el día deportivo) o excursiones
SC22Q06	Estudiantes que faltan el respeto a los docentes
SC22Q07	Estudiantes que interrumpen u obstaculizan el desarrollo normal de las clases
SC22Q08	Uso de alcohol o drogas ilegales por parte de los estudiantes
SC22Q09	Estudiantes que intimidan o amenazan a otros estudiantes
SC22Q10	Estudiantes que no están motivados a desarrollar todo su potencial
SC22Q11	Malas relaciones entre estudiantes y docentes
SC22Q12	Docentes que deben enseñar a estudiantes de distintos niveles de habilidad dentro de una misma clase
SC22Q13	Docentes que deben enseñar a estudiantes de diversos orígenes étnicos
SC22Q14	Docentes que tienen bajas expectativas sobre sus estudiantes
SC22Q15	Docentes que NO atienden las necesidades individuales de sus estudiantes
SC22Q16	Inasistencia por parte de los docentes
SC22Q17	Personal que se resiste a los cambios
SC22Q18	Docentes que son demasiado estrictos con los estudiantes
SC22Q19	Docentes que no llegan bien preparados a sus clases

Fuente: OCDE-PISA

4.2.-Resultados del Análisis EM No Métrico en SPSS

Tabla N° 4.- Medidas de Bondad Ajuste del Modelo

Iteration	S-stress	Improvement
1	0,15489	0,3108
2	0,12380	0,00200
3	0,12181	0,00033
4	0,12148	

Fuente: Elaboración Propia

El algoritmo iterativo tiende a minimizar el criterio de ajuste del modelo del S-stress (Tabla4.2.) el proceso se detuvo cuando el coeficientes de S-stress es menor que el valor de convergencia (0.001000). El proceso se detuvo en la cuarta interacción con el valor 0.12148, lo que significa que el ajuste del modelo es bueno.

Tabla N° 5 .Valor de Stress y RSQ

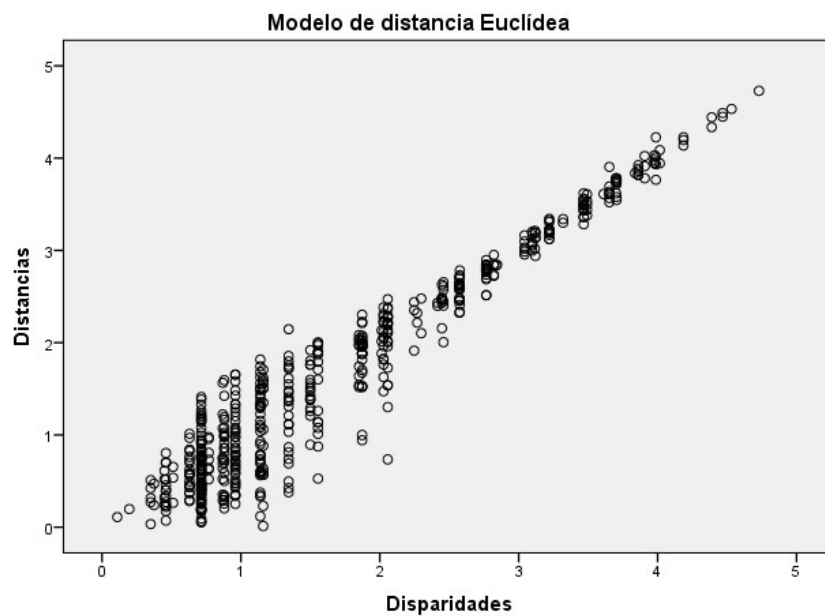
Stress	RSQ
0,15463	0,92630

Fuente: Elaboración Propia.

El valor del coeficiente al finalizar la estimación no lineal es Stress = 0,15463 y la proporción de variabilidad explicada, RSQ = 0,92630.

Stress y del RSQ no tenemos claro si el ajuste es bueno o no, también podemos analizar el gráfico de los residuos, entre las disparidades y las distancias, donde observamos un ajuste lineal creciente, que significa un buen ajuste y que el valor bajo del stress no es consecuencia de una no linealidad. (Figura N°4 y N°5)

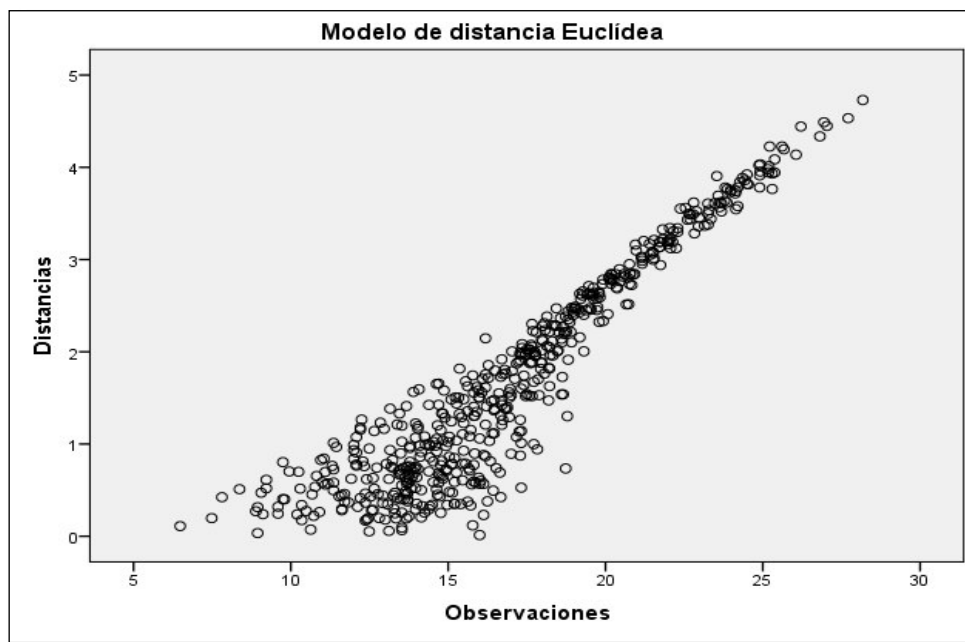
Figura N°4: Gráfico de Ajuste Lineal.



Fuente: Elaboración Propia.

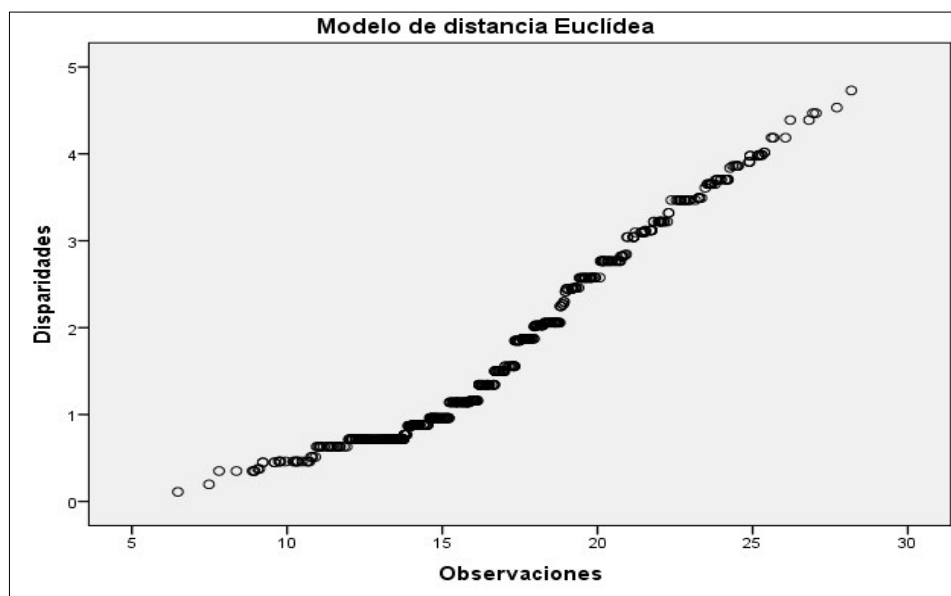
Como se observa en el gráfico anterior, las distancias euclídeas entre los puntos (Figura N°4), se aproximan con bastante exactitud a las similitudes entre parejas de las I.E. evaluadas, tampoco se observan outliers o puntos separados del gráfico de ajuste.

Figura N°5: Gráfico de Ajuste No Lineal.



Fuente: Elaboración Propia.

Figura N°6: Grafico Shepard



Fuente: Elaboración Propia.

El Figura N°6, se puede apreciar del diagrama de Shepard, complementar la medida del stress, con el diagrama de dispersión que representa a las proximidades frente a las disparidades, lo que nos permite detectar los “outliers” y la forma de nuestra una función creciente lo que se determina el buen ajuste la que representa en forma diagonal como se refleja en nuestro caso.

Tabla N° 6. Medidas de ajuste y stress

Prueba Estadística	Valor
Stress bruto normalizado	,01464
Stress-I	,12101 ^a
Stress-II	,22791 ^a
S-Stress	,02609 ^b
Dispersión explicada (D.A.F.)	,98536
Coefficiente de congruencia de Tucker	,99265

Fuente: Elaboración Propia.

Nota: PROXSCAL minimiza el stress bruto normalizado.

a. Factor para escalamiento óptimo = 1,015.

b. Factor para escalamiento óptimo = ,998.

De las medidas de ajustes y Stress ,el Stress bruto normalizado (0.01040) tiende a cero el cual indica un buen ajustes al que el Stess-I, Stress-II, ya que tienden a cero. La dispersión de explicada y el coeficiente de Tucker (0.9896 y 0.9947) tiende a 1, lo cual indica un mejor ajuste.

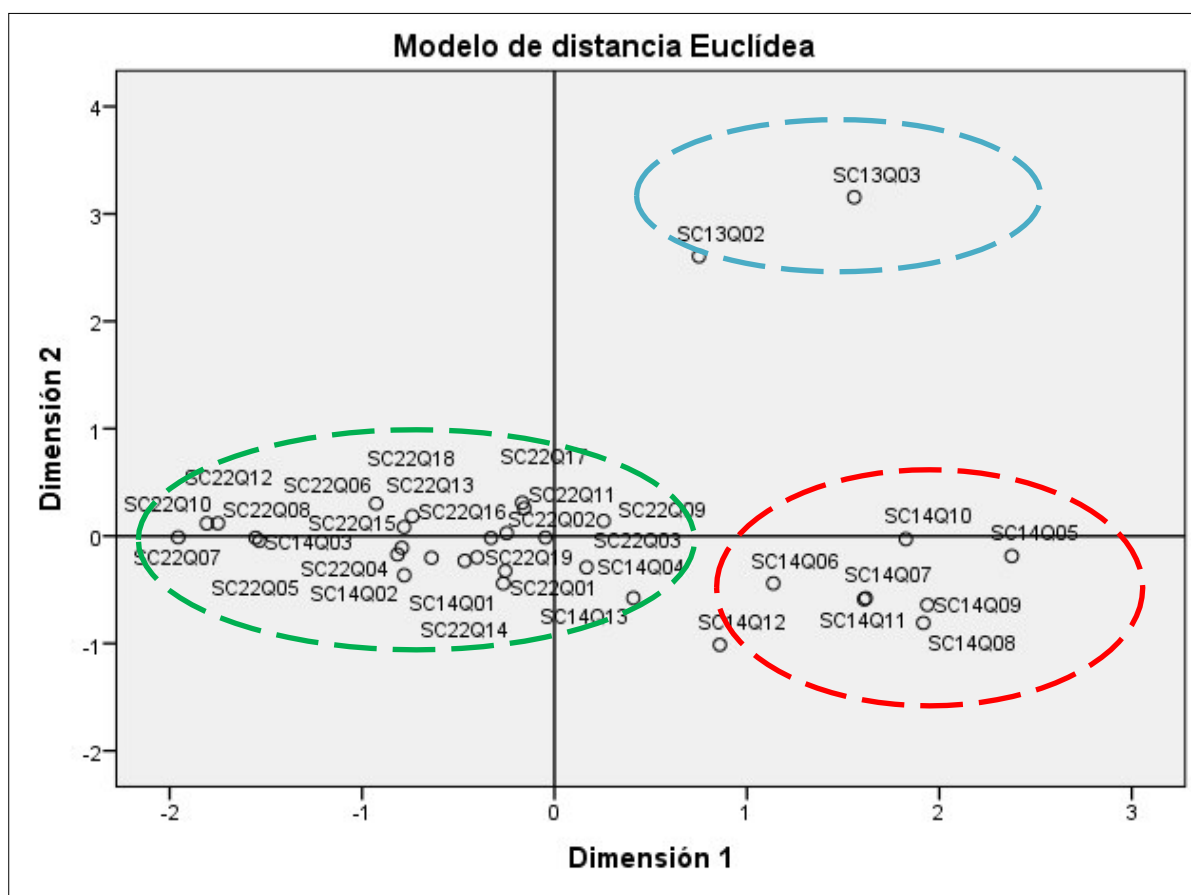
Las coordenadas obtenidas sobre los factores cada dimensión muestra en su posición respecto a los ejes (ver Tabla N° 7).

Tabla N° 7. Coordenadas en la Configuración
Bidimensional

Dimensión				
N°	Código	Variable	1	2
1	SC13Q03	Para trabajos o proyectos	1,5574	3,1548
2	SC13Q02	Para las tareas dejadas para la casa	0,7506	2,6048
3	SC14Q01	Falta de profesores de Ciencias calificados	-0,2664	-0,4419
4	SC14Q02	Falta de profesores de Matemática calificados	-0,781	-0,365
5	SC14Q03	Falta de profesores de Comunicación calificados	-0,816	-0,1718
6	SC14Q04	Falta de profesores calificados de otras asignaturas	0,1668	-0,2922
7	SC14Q05	Escasez o deficiencia de equipamiento del laboratorio de Ciencias	2,3764	-0,1867
8	SC14Q06	Escasez o deficiencia de material educativo (por ejemplo, libros de texto)	1,137	-0,4406
9	SC14Q07	Escasez o deficiencia de computadoras para la enseñanza	1,6178	-0,5789
10	SC14Q08	Falta o deficiencia de conexión a Internet	1,9168	-0,8091
11	SC14Q09	Escasez o deficiencia de software (programas de computadora) para la enseñanza	1,9378	-0,6406
12	SC14Q10	Escasez o deficiencia de materiales de biblioteca	1,826	-0,028
13	SC14Q11	Escasez o deficiencia de edificaciones y terrenos del colegio	1,6089	-0,5886
14	SC14Q12	Escasez o deficiencia de sistemas de enfriamiento o calentamiento y sistemas de electrificación	0,859	-1,0124
15	SC14Q13	Escasez o deficiencia de espacio para la enseñanza	0,4099	-0,578
16	SC22Q01	Ausentismo escolar no justificado	-0,2558	-0,3244
17	SC22Q02	Estudiantes que no asisten a algunas de las clases durante la jornada escolar	-0,3303	-0,018
18	SC22Q03	La dificultad de aprendizaje - Estudiantes llegando tarde	-0,0476	-0,0134
19	SC22Q04	Estudiantes que llegan tarde al colegio	-0,7935	-0,1099
20	SC22Q05	Estudiantes que no asisten a eventos escolares obligatorios (por ejemplo, el día deportivo) o excursiones	-1,5533	-0,0169
21	SC22Q06	Estudiantes que faltan el respeto a los docentes	-0,9272	0,3026
22	SC22Q07	Estudiantes que interrumpen u obstaculizan el desarrollo normal de las clases	-1,9556	-0,012
23	SC22Q08	Uso de alcohol o drogas ilegales por parte de los estudiantes	-1,5332	-0,0446
24	SC22Q09	Estudiantes que intimidan o amenazan a otros estudiantes	0,255	0,1404
25	SC22Q10	Estudiantes que no están motivados a desarrollar todo su potencial	-1,7493	0,119
26	SC22Q11	Malas relaciones entre estudiantes y docentes	-0,1575	0,2588
27	SC22Q12	Docentes que deben enseñar a estudiantes de distintos niveles de habilidad dentro de una misma clase	-1,8071	0,118
28	SC22Q13	Docentes que deben enseñar a estudiantes de diversos orígenes étnicos	-0,6392	-0,2028
29	SC22Q14	Docentes que tienen bajas expectativas sobre sus estudiantes	-0,4669	-0,2309
30	SC22Q15	Docentes que NO atienden las necesidades individuales de sus estudiantes	-0,7817	0,0834
31	SC22Q16	Inasistencia por parte de los docentes	-0,2486	0,0272
32	SC22Q17	Personal que se resiste a los cambios	-0,1675	0,3104
33	SC22Q18	Docentes que son demasiado estrictos con los estudiantes	-0,7394	0,1852
35	SC22Q19	Docentes que no llegan bien preparados a sus clases	-0,4026	-0,1981

Fuente: Elaboración Propia.

Figura N°7: Mapa Perceptual de Lo Factores de la I.E.



Fuente: Elaboración Propia.

Del mapa perceptual del Figura N°7; cómo podemos observar está conformada un sub grupo por las variables tareas dejados para la casa (SC13Q02) y la variable trabajos o proyectos (SC13Q03), son las que caracterizan a las Instituciones Educativas que fomentan los trabajos o proyectos en sus establecimientos y promoviendo la tareas para la casa de los estudiantes.

Se puede observar que las variables conformada por Escasez o deficiencia de equipamiento del laboratorio de Ciencias (SC14Q05), Escasez o deficiencia de material educativo (por ejemplo, libros de texto (SC14Q06), Escasez o deficiencia de computadoras para la enseñanza (SC14Q07), Falta o deficiencia de conexión a Internet(SC22Q08), y

Escasez o deficiencia de software (programas de computadora) para la enseñanza, Escasez o deficiencia de software (programas de computadora) para la enseñanza (SC14Q09), Escasez o deficiencia de materiales de biblioteca (SC14Q10), Escasez o deficiencia de edificaciones y terrenos del colegio(SC14Q11) y Escasez o deficiencia de sistemas de enfriamiento o calentamiento y sistemas de electrificación (SC14Q12), son las que caracterizan a las Instituciones Educativas por la falta y deficiencia en la infraestructura (laboratorios, biblioteca computadoras, etc.), materiales con que debería contar toda una institución educativa.

Del mapa perceptual resultante se puede establecer de las variables conformadas por las variables Falta de profesores de Ciencias calificados (SC14Q01), Falta de profesores de Matemática calificados (SC14Q02), Falta de profesores de Comunicación calificados(SC14Q03), Falta de profesores calificados de otras asignaturas(SC14Q04) , Estudiantes que no asisten a algunas de las clases durante la jornada escolar (SC22Q02), La dificultad de aprendizaje - Estudiantes llegando tarde(SC22Q03), Estudiantes que llegan tarde al colegio (SC22Q04), Estudiantes que no asisten a eventos escolares obligatorios (SC22Q05), Estudiantes que faltan el respeto a los docentes (SC22Q06), Estudiantes que interrumpen u obstaculizan el desarrollo normal de las clases (SC22Q07), Uso de alcohol o drogas ilegales por parte de los estudiantes (SC22Q08), Estudiantes que intimidan o amenazan a otros estudiantes (SC22Q09), Estudiantes que no están motivados a desarrollar todo su potencial (SC22Q10), Malas relaciones entre estudiantes y docentes (SC22Q11), Docentes que deben enseñar a estudiantes de distintos niveles de habilidad dentro de una misma clase (SC22Q12), Docentes que deben enseñar a estudiantes de diversos orígenes étnicos (SC22Q13), Docentes que tienen bajas expectativas sobre sus estudiantes (SC22Q14), Docentes que NO atienden las necesidades individuales de sus estudiantes (SC22Q15), Inasistencia por parte de los docentes (SC22Q16), Personal que se resiste a los cambios (SC22Q17), Docentes que son demasiado estrictos con los estudiantes (SC22Q17), Docentes que no llegan bien preparados a sus clases (SC22Q19), son las que caracterizan a las Instituciones Educativas por que presentan Falta de profesores calificados para las materias de ciencias, Matemática y comunicación, ausentismo por parte de los estudiantes y docentes , una falta de compromiso por parte de los principales actores que intervienen en el proceso de enseñanza.

CONCLUSIONES

De los resultados obtenidos (Mapa Perceptual), mediante la aplicación del Escalamiento multidimensional (EM), se analizó el posicionamiento de las variables de las instituciones Educativas que participaron Programa Internacional de Evaluación de Estudiantes (PISA), la cual se caracterizan:

Mediante la técnica de análisis de Escalamiento multidimensional (EM), se han agrupado a aquellas a las Instituciones Educativas en tres subgrupos; el primer grupo de instituciones educativas donde se fomenta los proyectos o trabajos de investigación así como el establecimiento de tareas para la casa, con lo cual se busca una interacción entre estudiantes y docentes, el segundo grupo formado por aquellas instituciones educativas se caracterizan por la falta y deficiencia en sus infraestructuras (laboratorios, biblioteca computadoras, etc.), herramientas principales con que debería contar toda Institución Educativa, el tercer grupo de Instituciones Educativas que se caracterizan por presentar y/o asentar falta de plana de Docentes (profesores) calificados para las materias de ciencias, Matemática y comunicación, ausentismo por parte de los estudiantes y docentes; un déficit de compromiso por parte de los principales actores que intervienen en este proceso de enseñanza.

BIBLIOGRAFÍA:

- Blanco, L. & Mujica, C. (1998) *Representación De Variables Sobre Una Configuración De Objetos Obtenida A Través De Un Escalamiento Multidimensional*. Revista Venezolana de Análisis de Coyuntura, 1998, Vol. IV, No. 2 (jul-dic), pp. 223-236.
- Corrales, J. (2006), publica un artículo: *Escalamiento Multidimensional Para Determinar Modelos De Gestión Administrativa*, Theoria, Vol. 15 (2): 17-31, Universidad de Bío-Bío-CHILE
- Cox Trevor F.& Cox Michael A.A. (2001) *Multidimensional Scaling*. Segunda Edición, Editorial Chapman & Hall / CRC
- Gower, & Legendre, P. (1986), *Metric and Euclidean Properties of Dissimilarity Coefficients*, Journal of Classification 3:5-48 (1986), Rothamsted Experimental Station-Université de Montréal-EE.UU:
- Guerrero, F., & Ramírez, J. (2010). *El Análisis de Escalamiento Multidimensional: una Alternativa y un complemento a otras Técnicas Multivariantes con una aplicación al sector Turístico de Andalucía*. International Journal of Environmental Analytical Chemistry.
- Hair & Anderson & Tatham & Black (1999) *Análisis Multivariante*. Quinta Edición Editorial Prentice Hall.
- <http://www.oecd.org/pisa/>. (2012). *pagina web. Retrieved from OCDE. (2012). Informe_PISA_2012_Peru.pdf.*
- http://umc.minedu.gob.pe/wp-content/uploads/2013/12/reporte_pisa_2012.pdf. (Diciembre de 2013). Obtenido de http://umc.minedu.gob.pe/wp-content/uploads/2013/12/reporte_pisa_2012.pdf
- http://www.perueduca.pe/foro//message_boards/message/53636512?_19_advancedSearch=false&_19_keywords=&_19_delta=20&_19_resetCur=false&_19_threadView=tree&_19_cur=1&_19_andOperator=true

Moral, J., Sánchez Juan, & Villareal María. (2010). *Desarrollo de una Escala Multidimensional Breve de Ajuste Escolar*. Revista Electrónica de Metodología Aplicada, 15(nº1), 1–11.

Porcar, M & Escalante, E. (2000). *Límites de la explicación interna (norma de internalidad). Análisis de escalamiento multidimensional*
<http://www.revistareid.net/revista/n2/REID2art4.pdf>

Rangel, C. (2012), *Evaluación de los síntomas emocionales dentro del constructo de calidad de vida en pacientes con cáncer empleando el escalamiento multidimensional*, www.bdigital.unal.edu.co. Universidad Nacional de Colombia-COLOMBIA.

Salazar, M. (2006). *Análisis Procrustes: Comparación del Escalamiento Multidimensional y Analisis de Correpondencia*. UNMSM-Lima.

Sánchez, J. & Villareal, M. (2010), presenta un estudio: *Desarrollo de una Escala Multidimensional Breve de Ajuste Escolar*, Revista Electrónica de Metodología Aplicada 2010, Vol. 15 nº1, pp. 1-11 Universidad Autónoma de Nuevo León-MEXICO.

Serralvo, F & Tadeu, M.(2005). *Tipologías del posicionamiento de marcas. Un estudio conceptual en Brasil y en España*. Revista Galega de Economía, vol. 14, núm. 1-2, junio-diciembre, 2005, Universidad de Santiago de Compostela -España.

Socha, A. & Sigler, E. (2012), *Using Multidimensional Scaling To Improve Functionality Of The Revised Learning Process Questionnaire*. Assessment & Evaluation in Higher Education .Vol. 37, No. 4, June 2012, 409–425. Western Carolina University -EE.UU.

Ucedo, V. (2013). *Comparación De Los Modelos Logit Y Probit Del Análisis Multinivel, En El Estudio Del Rendimiento Escolar*. UNMSM-Lima.

ANEXOS I

SALIDAS DEL ESCALAMIENTO MULTIDIMENSIONAL (SPSS)

Tabla N° I.1. Alscal Procedure Options

Alscal Procedure Options - Data Options-	
Number of Rows (Observations/Matrix).	34
Number of Columns (Variables) . . .	34
Number of Matrices	1
Measurement Level	Ordinal
Data Matrix Shape	Symmetric
Type	Dissimilarity
Approach to Ties	Leave Tied
Conditionality	Matrix
Data Cutoff at	,000000

Tabla N° I.2. Model Options- Model

Model Options- Model	
Model	Euclid
Maximum Dimensionality	2
Minimum Dimensionality	2
Negative Weights	Not Permitted

Tabla N° I.3. Output Options

Output Options-	
Job Option Header	Printed Data
Matrices	Printed
Configurations and Transformations .	Plotted
Output Dataset	Not Created
Initial Stimulus Coordinates . . .	Computed

Tabla N° I.4. Output Options

Algorithmic Options-	
Maximum Iterations	30
Convergence Criterion	,00100
Minimum S-stress	,00500
Missing Data Estimated by	Ulbounds
Tiestore	561

Tabla N° I.5. Raw (unscaled) Data for Subject 1

	1	2	3	4	5	6	7	8	9	10
1	,000									
2	13,784	,000								
3	24,920	22,023	,000							
4	25,612	22,045	9,220	,000						
5	25,377	21,726	9,220	7,483	,000					
6	24,125	21,166	11,705	12,490	12,000	,000				
7	23,345	22,068	19,287	20,952	21,749	17,804	,000			
8	23,643	21,517	15,811	16,703	17,464	14,595	12,247	,000		
9	23,979	22,293	17,664	20,075	19,053	16,462	14,422	14,142	,000	
10	25,159	23,259	18,708	20,761	20,125	18,193	15,811	15,033	13,711	,000
11	24,536	22,978	18,520	19,900	20,199	16,793	12,610	13,229	14,177	12,369
12	21,954	20,928	17,972	19,183	19,799	15,556	11,180	9,747	15,199	15,460
13	24,145	22,159	17,944	19,313	19,157	18,193	16,186	16,432	16,000	16,248
14	25,219	22,804	17,292	18,055	17,607	17,720	18,628	15,969	17,292	17,176
15	23,537	21,213	15,264	16,186	15,427	15,620	19,313	16,523	16,703	17,578
16	24,495	20,976	15,780	17,321	15,875	16,673	19,621	16,882	17,234	17,692
17	23,580	20,199	14,731	16,125	14,629	16,000	19,468	17,000	17,578	18,138
18	22,383	20,174	15,033	16,340	15,133	15,199	19,183	16,733	16,733	18,000
19	24,880	22,068	14,832	14,248	13,528	14,933	21,401	17,776	19,545	20,100
20	26,211	22,782	15,620	15,199	13,675	17,407	25,298	20,833	21,817	22,539
21	24,207	20,785	14,387	14,697	13,638	15,100	22,293	18,682	19,621	21,424
22	28,178	23,875	15,780	15,492	14,765	18,601	26,814	22,249	23,854	24,920
23	27,037	22,650	15,362	14,799	14,036	16,279	24,900	20,347	22,000	23,281
24	22,825	20,664	15,100	15,716	16,462	12,923	16,186	14,629	17,493	18,815
25	26,944	22,891	15,969	15,684	14,422	18,330	26,058	21,749	22,694	23,812
26	23,152	20,736	14,933	16,125	15,556	14,000	19,209	16,703	18,248	19,925
27	27,713	23,622	17,407	17,263	16,310	18,493	25,671	21,517	23,259	24,269
28	25,199	21,703	13,266	13,748	13,000	14,526	21,166	17,029	18,385	19,799
29	25,179	21,494	13,964	14,491	14,000	13,638	20,174	15,906	18,303	19,723
30	24,434	20,761	13,928	14,663	13,528	15,000	22,000	17,607	19,026	20,785
31	23,685	20,469	13,711	14,731	13,675	13,964	19,494	16,371	17,776	18,868
32	21,817	18,439	13,748	14,142	13,784	13,928	19,824	15,969	17,748	18,574
33	23,979	20,322	13,038	13,528	12,369	14,526	21,726	17,378	18,921	20,881
34	24,352	21,142	12,570	13,379	12,207	13,675	20,149	16,062	17,664	18,974

	11	12	13	14	15	16	17	18	19	20
11	,000									
12	11,916	,000								
13	15,716	16,031	,000							
14	17,321	16,852	14,107	,000						
15	17,889	17,664	13,528	13,784	,000					
16	18,762	18,921	17,635	16,673	15,684	,000				
17	18,815	19,183	17,578	16,613	15,492	8,944	,000			
18	17,349	17,176	16,186	15,906	14,933	11,705	11,619	,000		
19	19,925	19,416	19,391	17,972	16,941	13,748	13,304	13,191	,000	
20	24,145	23,259	21,541	19,261	17,349	14,799	13,964	15,232	12,083	,000
21	21,541	20,248	20,372	18,221	17,321	14,629	13,491	13,153	11,533	9,950
22	25,377	24,900	23,473	21,166	18,974	16,673	15,620	17,292	14,036	9,747
23	23,685	22,956	22,136	19,723	17,972	15,843	14,663	15,492	12,728	8,944
24	16,279	14,866	18,601	18,788	18,735	16,643	15,460	14,967	15,033	17,944
25	25,298	24,207	22,605	20,591	18,655	15,811	14,697	16,093	14,107	9,110
26	18,000	17,029	17,748	18,221	17,321	15,811	14,422	14,387	15,652	16,703
27	24,495	23,791	22,694	20,445	19,799	16,186	15,100	16,823	15,264	12,610
28	19,570	19,000	18,439	17,861	16,401	13,601	12,923	13,115	12,410	12,000
29	18,815	17,664	18,628	18,655	17,833	16,125	14,765	15,133	15,395	15,460
30	20,567	19,621	19,748	17,292	16,462	13,748	12,369	13,565	13,342	12,083
31	18,303	17,635	17,776	17,464	17,000	15,199	13,528	14,142	15,748	15,811
32	18,055	17,493	17,292	17,720	16,000	13,856	11,832	13,077	13,964	14,387
33	20,712	19,621	19,494	17,748	16,462	14,387	12,767	13,491	13,266	12,000
34	18,735	18,083	18,221	16,941	15,843	13,675	12,450	13,711	13,491	12,961

	21	22	23	24	25	26	27	28	29	30
21	,000									
22	12,083	,000								
23	10,247	7,810	,000							
24	14,036	18,894	16,492	,000						
25	11,045	9,592	8,888	18,466	,000					
26	14,765	18,221	16,217	12,530	16,971	,000				
27	13,491	12,806	12,124	18,138	13,115	14,697	,000			
28	11,358	13,454	12,083	13,784	12,207	13,454	13,379	,000		
29	13,416	16,000	14,248	12,288	14,832	14,697	16,793	10,344	,000	
30	10,909	12,207	11,314	14,832	11,446	14,799	13,528	9,592	11,619	,000
31	13,820	17,117	14,900	12,649	15,330	13,304	16,941	10,677	10,344	10,770
32	12,329	15,362	13,675	13,675	14,071	12,490	14,629	11,091	12,410	10,817
33	10,724	12,845	10,954	14,283	11,358	14,526	13,964	9,798	11,358	6,481
34	11,358	13,892	12,649	13,266	13,153	13,820	15,264	10,198	10,630	9,055
	31	32	33	34						
31	,000									
32	11,619	,000								
33	10,296	11,000	,000							
34	10,488	11,091	8,367	,000						

Tabla N° I.6. Iteration history for the 2 dimensional solution (in squared distances)

Young's S-stress formula 1 is used.		
Iteration	S-stress	Improvement
1	,15489	
2	,12380	,03108
3	,12181	,00200
4	,12148	,00033

Iterations stopped because

S-stress improvement is less than ,001000

Stress and squared correlation (RSQ) in distances

RSQ values are the proportion of variance of the scaled data (disparities)

in the partition (row, matrix, or entire data) which

is accounted for by their corresponding distances.

Tabla N° I.7.. Stress values are Kruskal's stress formula 1.

For matrix	
Stress = ,15463	RSQ = ,92630

Tabla N° I.8. Coordenadas de en la Configuración
Bidimensional

Stimulus Coordinates			
Dimension			
Stimulus	Stimulus	1	2
Number	Name		
1	SC13Q03	1,5574	3,1548
2	SC13Q02	,7506	2,6048
3	SC14Q01	-,2664	-,4419
4	SC14Q02	-,7810	-,3650
5	SC14Q03	-,8160	-,1718
6	SC14Q04	,1668	-,2922
7	SC14Q05	2,3764	-,1867
8	SC14Q06	1,1370	-,4406
9	SC14Q07	1,6178	-,5789
10	SC14Q08	1,9168	-,8091
11	SC14Q09	1,9378	-,6406
12	SC14Q10	1,8260	-,0280
13	SC14Q11	1,6089	-,5886
14	SC14Q12	,8590	-1,0124
15	SC14Q13	,4099	-,5780
16	SC22Q01	-,2558	-,3244
17	SC22Q02	-,3303	-,0180
18	SC22Q03	-,0476	-,0134
19	SC22Q04	-,7935	-,1099
20	SC22Q05	-1,5533	-,0169
21	SC22Q06	-,9272	,3026
22	SC22Q07	-1,9556	-,0120
23	SC22Q08	-1,5332	-,0446
24	SC22Q09	,2550	,1404
25	SC22Q10	-1,7493	,1190
26	SC22Q11	-,1575	,2588
27	SC22Q12	-1,8071	,1180
28	SC22Q13	-,6392	-,2028
29	SC22Q14	-,4669	-,2309
30	SC22Q15	-,7817	,0834
31	SC22Q16	-,2486	,0272
32	SC22Q17	-,1675	,3104
33	SC22Q18	-,7394	,1852
34	SC22Q19	-,4026	-,1981

ANEXOS II

Tabla N°II.1 Distribución de Institución Educativa por Tipo de Gestión

Gestión	N° Instituciones Educativas	%	N° Estudiantes	%
Estatál	1 0	75,00%	4700	77,90%
No Estatal	60	25,00%	1335	22,10%

Fuente: UMC. Base de datos PISA 2012

Tabla N° II.2 Distribución de la Institución Educativa por Área geográfica

Área	N° Instituciones Educativas	%	N° Estudiantes	%
Urbano	1 99	82,90%	5461	90,50%
Rural	41	17,10%	574	9,50%

Fuente: UMC. Base de datos PISA 2012

Tabla N° II.3 Distribución de la Institución Educativa por Nivel y/o Modalidad

Nivel y/o Modalidad	N° Instituciones Educativas	%	N° Estudiantes	%
CEBA	6	2,50%	88	1,50%
Educación a Distancia	1	0,40%	5	0,10%
Regular	227	94,60%	5888	97,60%
Secundaria para Adultos	6	2,50%	54	0,90%

Fuente: UMC. Base de datos PISA 2012

Tabla N° II.4 Distribución de la muestra evaluada de Institución Educativa y Estudiantes
según Departamento

Departamento	N° Instituciones Educativas	%	N° Estudiantes	%
Amazonas	4	1,7%	105	1,7%
Ancash	10	4,2%	227	3,8%
Apurímac	4	1,7%	92	1,5%
Arequipa	12	5,0%	322	5,3%
Ayacucho	8	3,3%	169	2,8%
Cajamarca	13	5,4%	263	4,4%
Callao	7	2,9%	176	2,9%
Cusco	12	5,0%	313	5,2%
Huancavelica	4	1,7%	81	1,3%
Huánuco	7	2,9%	150	2,5%
Ica	6	2,5%	188	3,1%
Junín	11	4,6%	254	4,2%
La Libertad	15	6,2%	366	6,1%
Lambayeque	11	4,6%	285	4,7%
Lima	62	25,8%	1684	27,9%
Loreto	7	2,9%	155	2,6%
Madre De Dios	1	0,4%	17	0,3%
Moquegua	2	0,8%	63	1,0%
Pasco	3	1,2%	69	1,1%
Piura	15	6,2%	404	6,7%
Puno	11	4,6%	287	4,8%
San Martín	6	2,5%	147	2,4%
Tacna	4	1,7%	85	1,4%
Tumbes	1	0,4%	33	0,5%
Ucayali	4	1,7%	100	1,7%
Total	240	100,0%	6035	100,0%

Fuente: UMC. Base de datos PISA 2012